REPORT DOCUMENTATION PAGE

Form Approved
OMB NO. 0704-0188

1. AGENCY USE ONLY (Leave Blank)

2. REPORT DATE August 1958

3. REPORT TYPE AND DATES COVERED

4. TITLE AND SUBTITLE
Proceedings of the Third Conference on the Design of Experiments in Army Research, Development and Testing

5. FUNDING NUMBERS

6. AUTHOR(S)
Not Available

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Army Mathematics Advisory Panel

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)

U. S. Army Research Office
P.O. Box 12211
Research Triangle Park, NC 27709-2211

10. SPONSORING / MONITORING AGENCY REPORT NUMBER

ARO-OORR 58-5

11. SUPPLEMENTARY NOTES
The views, opinions and/or findings contained in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy or decision, unless so designated by other documentation.

12 a. DISTRIBUTION / AVAILABILITY STATEMENT

Approved for public release; distribution unlimited.

12 b. DISTRIBUTION CODE

13. ABSTRACT (Maximum 200 words)

This is a Technical report resulting from the Proceedings of the Third Conference on the Design of Experiments in Army Research, Development and Testing.

14. SUBJECT TERMS

15. NUMBER OF PAGES
~~273~~ 363

16. PRICE CODE

17. SECURITY CLASSIFICATION OR REPORT	18. SECURITY CLASSIFICATION ON THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
UNCLASSIFIED	**UNCLASSIFIED**	**UNCLASSIFIED**	**UL**

NSN 7540-01-280-5500

Standard Form 298 (Rev.2-89)
Prescribed by ANSI Std. 239-18
298-102

OORR 58-5

Office of Ordnance Research

PROCEEDINGS OF THE THIRD CONFERENCE ON THE DESIGN OF EXPERIMENTS IN ARMY RESEARCH DEVELOPMENT AND TESTING

OFFICE OF ORDNANCE RESEARCH, U. S. ARMY
BOX CM, DUKE STATION
DURHAM, NORTH CAROLINA

OFFICE OF ORDNANCE RESEARCH
Report No. 58-5
August 1958

PROCEEDINGS OF THE THIRD CONFERENCE ON THE DESIGN OF EXPERIMENTS IN ARMY RESEARCH DEVELOPMENT AND TESTING

Sponsored by the Army Mathematics Steering Committee
conducted at
Diamond Ordnance Fuze Laboratories
and
National Bureau of Standards
16-18 October 1957

OFFICE OF ORDNANCE RESEARCH, U. S. ARMY
BOX CM, DUKE STATION
DURHAM, NORTH CAROLINA

Initial Distribution

The initial distribution list of the Proceedings of the Third Conference on the Design of Experiments in Army Research, Development and Testing includes those who attended the meeting and/or the government installations with which they are associated. For economy, only a limited number of copies have been sent to each.

TABLE OF CONTENTS

---

*This paper was presented at the Conference. It is not published in these proceedings.

**On the program this paper appears under the joint authorship of P. G. Sanders and Boyd Harshbarger.

***This paper is to be issued in a classified security information (Confidential) appendix of this Technical Manual.

TABLE OF CONTENTS (Cont'd)

*This paper was presented at the Conference. It is not published in these proceedings.

---

*This paper was presented at the Conference. It is not published in these proceedings

**This paper is to be issued in a classified security information (Confidential) appendix of this Technical Manual.

FOREWORD

The Army Mathematics Advisory Panel, now called the Army Mathematics Steering Committee (AMSC), was established in 1954 by the Office of Ordnance Research to provide advice on the mathematical needs of the Army to the Chief, Research and Development, Office, Deputy Chief of Staff for Plans and Research, Department of the Army. Soon after it was organized the AMSC conducted a survey of the mathematical activities and requirements of more than 30 Army research, development and testing facilities. One of the most frequently mentioned needs expressed by the scientific personnel of these establishments was for greater knowledge of the modern statistical theory of the design and analysis of experiments.

On the basis of this expression of interest the AMSC decided to sponsor an Army-wide conference of the design of experiments. To meet the needs of the various participating groups, three kinds of sessions were placed on the agenda. The first type of session consisted of invited papers by well-known authorities on the philosophy and general principles of the design of experiments. The second type consisted of technical papers contributed by research workers from various Army research and development facilities. The third type, called clinical sessions, consisted of presentations and discussions of partially solved and unsolved problems arising in these installations. The success of the first meeting which was held October of 1955 has led to subsequent meetings being organized along similar lines.

To date three in this series of conferences have been conducted. All of them have been held at The Diamond Ordnance Fuze Laboratories and the National Bureau of Standards. The Third Conference held 16-18 October 1957 was attended by 203 registrants and participants from 69 organizations. Speakers and other participants came from Air Force Chief of Staff for Intelligence, Bell Telephone Laboratories, Bureau of Ordnance, U. S. Navy, General Electric Company, Iowa State College, National Bureau of Standards, Princeton University, Rand Corporation, University of Cambridge, University of Georgia, Virginia Polytechnic Institute, Wayne State University, Wright Air Development Center, 17 Army facilities.

The present volume is the Proceedings of the Third Conference, and it contains 24 papers which were presented at this meeting. Three additional papers also presented at this meeting will be printed in a classified appendix. The papers are being made available in the present form in order to encourage a wider use of modern statistical principles of the design of experiments in research, development, and testing work of concern to the Army.

The members of the Army Mathematics Steering Committee take this opportunity to express their thanks to those research workers in the various Army research, development, and testing facilities who participated in the Conference; to Lt. Colonel J. A. Ulrich, the Commanding Officer of the Diamond Ordnance Fuze Laboratories and to Dr. A. V. Astin, the Director of the National Bureau of Standards, for making available the excellent facilities of their two organizations for the Conference; to Mr. John A. Wheeler who handled the details of the local arrangements for the Conference at both installations; and to Dr. F. G. Dressel of the Office of Ordnance Research who carried through the details involved in organizing the conference and in preparing the Proceedings.

Finally, the Chairman wishes to express his appreciation to his Advisory Committee, William G. Cochran, Churchill Eisenhart, Frank E. Grubbs, and Clifford J. Maloney for their help in organizing the Conference.

S. S. Wilks
Professor of Mathematics
Princeton University

THIRD CONFERENCE ON THE DESIGN OF EXPERIMENTS IN ARMY RESEARCH DEVELOPMENT AND TESTING

16 - 18 October 1957
Diamond Ordnance Fuze Laboratories
and
National Bureau of Standards

16 October 1957

REGISTRATION: 0900 - 0930 (Eastern Daylight Saving Time)

MORNING SESSION: 0930 - 1215 East Building Lecture Room
National Bureau of Standards

Chairman: Professor S. S. Wilks
Princeton University

Introductory Remarks: Mr. Maurice Apstein, Associate Technical Director of the Diamond Ordnance Fuze Laboratories

Practical Problems in Experimental Design
Sir Ronald A. Fisher, University of Cambridge

Experimentation by Simulation and Monte Carlo
Dr. A. W. Marshall, Rand Corporation

LUNCH: 1215 - 1345

There will be three Technical Sessions conducted Wednesday afternoon. The security classification for Session III is Secret. No clearances will be required for Sessions I and II.

TECHNICAL SESSION I: 1345 - 1615 - East Building Lecture Room,
National Bureau of Standards

Chairman: Mr. J. F. O'Neil
Springfield Armory

A Computer Control System for the Simulation of Aerodynamic Heating of Structures
J. T. Sawyer, Army Ballistic Missile Agency

Development of an Impact Test Apparatus for Materials in Contact with Liquid Oxygen
W. R. Lucas and W. A. Riehl, Army Ballistic Missile Agency

Experimental Investigation of the Responses of a Liquid in an Oscillating Container. Two parts: 1) Analytical Consideration. 2) Design and Performance of Experiments
H. F. Beduerftig and W. R. Eulitz, Army Ballistic Missile Agency

TECHNICAL SESSION II: 1345 - 1615 - Chemistry Building Lecture Room
National Bureau of Standards

Chairman: L. S. Gephart
Office of Ordnance Research

A Statistical Design to Estimate Parameters Affecting Velocity in a Gun-Ammunition System
Abraham Rosenfeld, Weapon Systems Laboratory

An Investigation of Test Instruments
Lt. E. L. Bombara and Boyd Harshbarger, Redstone Arsenal

The Analysis of Test Data for Purpose of Setting Specification Limits
P. G. Sanders and Boyd Harshbarger, Redstone Arsenal

TECHNICAL SESSION III: 1345 - 1615 - Avenue Annex Conference Room
Diamond Ordnance Fuze Laboratories

Security Classification - SECRET

Chairman: J. F. Sullivan
Watertown Arsenal

Techniques of Weapon Effectiveness Analysis
L. F. Nichols, Picatinny Arsenal

Analysis of Wind Speed Frequency Distributions and Their Application
H. G. Baussus, Army Ballistic Missile Agency

Experimental Investigation of the Motion of a Liquid in a Decelerated Container
E. A. Hellebrand, Army Ballistic Missile Agency

SOCIAL MIXER: 1730 - (Sheraton Park Hotel)

17 October 1957

Technical Sessions IV, V, VI will run concurrently on Thursday morning. All the papers in these Sessions are unclassified.

Thursday afternoon will be devoted to three Clinical Sessions. The papers in Sessions A and B carry no security classification. The three papers in Session C are classified Confidential.

TECHNICAL SESSION IV: 0900 - 1130 - Chemistry Building Lecture Room,
National Bureau of Standards

Chairman: A. C. Cohen, Jr.
University of Georgia

An Example of Automation with Associated Statistical Problems
E. L. Cox and W. D. Foster, Army Chemical Corps

TECHNICAL SESSION IV (Cont'd)

Design of an Experiment to Study the Effect of Balloon Size on Its Response to the Wind
Raymond Bellucci, Signal Corps Engineering Laboratories

Evaulation of Virus Preparations as to Potency
F. M. Wadley, Army Chemical Corps

Linear Structural Relationships Underlying the Decomposition of Levinstein H
Henry Ellner, Army Chemical Corps

TECHNICAL SESSION V: 0900 - 1130 - East Building Lecture Room
National Bureau of Standards

Chairman: Walter Pressman
Signal Corps Engineering Laboratories

Experimental Design in Field Studies on Leadership
Carl Lange and F. H. Palmer, Human Resources Research Office

The Design of Controlled Field Experiments
F. I. Hill, Technical Operations, Inc.

Determining Life Behavior of Sub-Miniature Tubes Through Designed Experiments
E. G. Bianco, General Electric Company

TECHNICAL SESSION VI: 0900 - 1130 - Materials Testing Laboratory Lecture Room, National Bureau of Standards

Chairman: Don Mittleman
Diamond Ordnance Fuze Laboratories

A Point of View in the Analysis of Simulation Data
Sol Haberman, Operations Research Office

Ultrasonics, a Tool for Weldment Inspection
W. R. Lucas and T. E. Kingsbury, Army Ballistic Missile Agency

Short Life Study of Capacitors
R. W. Tucker, Diamond Ordnance Fuze Laboratories

LUNCH: 1130 - 1315

CLINICAL SESSION A: 1315 - 1615 - East Building Lecture Room
National Bureau of Standards

Chairman: P. A. Rider
Wright Air Development Center

CLINICAL SESSION A (Cont'd)

Panel Members: G. E. P. Box, Princeton University
H. O. Hartley, Iowa State College
A. W. Marshall, Rand Corporation
John Tukey, Princeton University and Bell Telephone Laboratories
Joseph Weinstein, Signal Corps Engineering Laboratories

Problem for Estimating Tolerance Bands for Sample Curves from a Weiner Process
B. M. Kurkjian, Diamond Ordnance Fuze Laboratories

The Design of Control Simulation Experiments
M. D. Springer, Combat Operations Research Group

Problems in Analysis of Electron Tube Experiments
Mortimer Zinn, Signal Corps Engineering Laboratories

Electronic Circuitry Design Via Computer Simulation
R. Lacy, Signal Corps Engineering Laboratories

CLINICAL SESSION B: 1315 - 1615 - Chemistry Building Lecture Room, NBS

Chairman: A. Golub
Weapon Systems Laboratory

Panel Members: Churchill Eisenhart, National Bureau of Standards
Benjamin Epstein, Wayne State University
C. J. Maloney, Army Chemical Corps
William Pabst, Bureau of Ordnance, U. S. Navy

Correlation of Fuze Functioning with Detonator Static Sensitivity Tests
Benjamin Shratter, Lake City Arsenal

Some Problems Encountered in the Evaluation of Erosion in Cannon Bores
P. J. Loatman, Watervliet Arsenal

Determining Durability of Textile Fabrics by Means of Controlled Field Testing
J. W. Griswold, Quartermaster Research and Engineering Field Evaluation Agency

CLINICAL SESSION C: 1315 - 1615 - Avenue Annex Conference Room, DOFL

Chairman: O. P. Bruno
Weapons Systems Laboratory

Panel Memebers: Besse Day, Bureau of Ships, U. S. Navy
F. E. Grubbs, Weapon Systems Laboratories

CLINICAL SESSION C (Cont'd)

Panel Members: Joseph Weinstein, Signal Corps Engineering Laboratories
S. S. Wilks, Princeton University

Security Classification - The papers by R. H. Burros, J. H. Campagna, and A. Bulfinch carry a classification of CONFIDENTIAL.

Some Problems in the Quantitative Analysis of Combat Intelligence
R. H. Burros, Combat Operations Research Group

Testing Philosophy for Guided Missile Fuzes
J. H. Campagna, Diamond Ordnance Fuze Laboratories

Statistical Design for Experiment on Sensitivity of Explosives to Setback Pressures
A. Bulfinch, Picatinny Arsenal

18 October 1957

From 0900 - 1015 there will be three sessions conducted concurrently. Session IX contains two contributed papers which did not appear on the agenda issued along with the invitations to the conference.

The final phase of the conference will consist of two invited addresses that will be delivered in the East Building Conference Room from 1030 to 1300.

SESSION VII: 0900 - 1015 - Exhibit Hall, Industrial Building
National Bureau of Standards

Chairman: A. Bulfinch
Picatinny Arsenal

On the Relation Between the Engineer and the Statistician
Joseph Mandelson, Army Chemical Corps

SESSION VIII: 0900 - 1015 - Chemistry Building Lecture Room
National Bureau of Standards

Chairman: J. A. Greenwood
Air Force Chief of Staff for Intelligence

Design of an Experiment in the Reliability Analysis of a Complex Component
J. W. Mitchell, Frankford Arsenal

Manual on Experimental Statistics for Ordnance Engineers
Mary G. Natrella, National Bureau of Standards

SESSION IX: 0900 - 1015 - Materials Testing Laboratory Lecture Room, National Bureau of Standards

Chairman: E. L. Cox,
Army Chemical Corps

Punch Card Computing of F-Tests
G. H. Andrews, J. Dominitz, G. T. Eccles, C. J. Maloney, and C. W. Riggs, Army Chemical Corps

Compounding Confidence Regions
L. M. Court, Diamond Ordnance Fuze Laboratories

GENERAL SESSION: 1030 - 1300 - East Building Conference Room, National Bureau of Standards

Chairman: Colonel G. F. Leist, Ordnance Corps
Commanding Officer of the Office of Ordnance Research

Life Testing
Professor Benjamin Epstein, Wayne State University

Changes in the Outlook of Statistics Brought About by Modern Computers
Professor H. O. Hartley, Iowa State College

EXPERIMENTATION BY SIMULATION AND MONTE CARLO

A. W. Marshall
The Rand Corporation

I. INTRODUCTION. The title of this paper emphasizes one aspect of Monte Carlo and Simulation techniques, that is, that they are experimental techniques. Since both techniques involve experimentation one would suppose that they would make use of ideas drawn from the field of the statistical design of experiments. This has proved to be the case. However, Monte Carlo and Simulation involve only synthetic experiments, often involving only paper and pencils or computing machines. This is both a gain and a loss. From the design of experiment point of view their synthetic character generates an additional, and very flexible, degree of freedom. This added degree of freedom can be exploited, sometimes to an extraordinary extent, to increase the accuracy of the estimates obtained from the experiments. Reduction of the variance of estimates, through proper design, of several thousand-fold is not uncommon in Monte Carlo problems. On the other hand, the degree of approximation to reality of the models used in these synthetic experiments is often not known.

Simulation and Monte Carlo often go together. Indeed a common usage "Monte Carlo Simulation" tends to emphasize Monte Carlo as a simulation technique for problems having a probabilistic basis. However, Monte Carlo need have nothing to do with Simulation and, from a design of experiment point of view, the two techniques are usually antithetical.

Monte Carlo, at least as a major numerical analysis technique, was developed and pushed in the late 40's by von Neumann and Ulam. Its original development was for use in the solution of engineering and physics problems. Typical problems in these areas today have the following characteristics:

1. The problems are well formulated mathematically, but are often beyond solution by analytic or usual computational methods.

2. The problems are specific and require specific answers. One is not just looking for good ideas or interesting information.

3. The problems are such that the use of straightforward sampling would be very costly, because of the accuracy required in the answer.

These characteristics have had a great influence on the initial development of Monte Carlo; almost from the very beginning sophisticated sampling designs were investigated. Quite soon splitting, Russian Roulette, importance sampling, correlated sampling, etc were discovered, or rediscovered, as sampling techniques and applied to particle diffusion, shielding, and nuclear reactor problems. In applying Monte Carlo to their problems the physicists often based their models on the physical process back of their mathematical formulations. However they were not at all interested in the model as a simulation *per se*. Distortions of the model or its parameters were quickly introduced in order to reduce the variance

of the estimates.

The use of Monte Carlo is now widespread in areas other than those of physics and engineering. In particular operations analysis studies often use Monte Carlo. Its popularity in this type of study, where many of the characteristics of the problems are almost exactly opposite to those of physics and engineering, is interesting. Also the use of variance reducing techniques is seldom seen in operations analysis applications of Monte Carlo. This requires some explanation.

The basic reason for the extensive use of Monte Carlo in operations analysis problems is that it is the easiest computational method to apply to the very large and complicated models now typical of problems studied in that area. These problems often have prominent stochastic elements in them. They are also new problems. Their mathematical formulation preliminary to the application of more traditional methods of mathematical or numerical analysis would often require much effort. Once formulated the problems almost never have known analytic solutions. The application of the traditional methods of numerical analysis are difficult, if not impossible. In order to apply Monte Carlo methods, however, it is only necessary to be able to model the physical process to be studied. Since large, high-speed computing machines are available to take over the laborious part of these calculations, the use of Monte Carlo allows one to substitute brute force computational methods for mathematical ingenuity and thought. Indeed for a good many of the problems studied by operations analysts there is no feasible alternative to Monte Carlo. This is especially so if information on the probability distribution of the outcome of some process is required as well as information about the expected value. Traditional methods of analysis are completely hopeless in this case if the problem is at all complicated.

The explanation of the neglect of variance reducing techniques in operations analysis applications is more complicated. First, to some extent the analysts have not been sufficiently aware of the various possibilities. Second, in most operations analysis problems extremely accurate answers are not required. Therefore sample sizes are not very large even when purely random sampling is used. The lower accuracy requirements result from the fact that the models, no matter how complicated, may not be adequate and from the fact that the basic parameters in the problem are often only very uncertainly known. In addition the analyst is usually looking for, and only interested in, rather big differences between, for example, the currently used system and proposed improved systems. Third, there is often a great deal of interest in the simulation provided by a model and in the sampling of its operation that would be lost if some of the variance reducing techniques that distort reality were applied. Some of the reasons for this interest in realistic simulation are:

1. The belief that detailed observation of the process under study will lead to understanding it, and to suggestions for improving it not otherwise obtainable. There is indeed a good deal of evidence that human beings have a great capacity for understanding the workings of complicated systems, and can find near optimum decision rules, operating procedures, etc., if they have enough experience with the

system and it is stable enough.

2. The very great realism available through Monte Carlo Simulation makes it possible, it is thought, to sell the results of the study more easily. There should be fewer arguments about the adequacy of the model. And if the working of the model is visible, that is, not completely buried inside some computing machine, the client may by seeing the operation and outcome with his own eyes convince himself of the results.

3. Criteria for ranking one system over another, one of the knottiest of all problems in operations analysis, need not be specified in advance. This means that after observations have been made, very complicated valuations of performance can be used and something nearer to the true state of the interested parties' value preferences applied. The availability of concrete outcomes to be thought about and ranked may also facilitate cooperation between researcher and client in finding appropriate ranking criteria. Since the researcher may not have a good feeling for the value preferences of the client, improved cooperation on this problem may be very important.

For many people in the operations research area these advantages of simulation seem overwhelming. In many cases, these advantages do not seem so to me, given the fact that simulation can only be emphasized at the expense of variance reduction. Moreover there is a tendency to underestimate seriously the computing cost, even with high-speed machines, of carrying out simulation studies. There are two sources of error that produce this tendency. First, the cost of the initially planned computing job tends to be slightly underestimated for reasons well known to all who are familiar with machine computations. Second, and more important, the number of cases (parameter variations) usually run in the end exceeds many-fold the original estimate. Once a problem has been coded in one form and is running it is almost never recoded to incorporate new design features. Therefore, what seemed like a small price to pay for the advantages of simulation is multiplied many-fold, perhaps by factors of 20 or 50, or even 100. More forethought as to the real cost of straightforward sampling is in order, especially since some variance reducing techniques are not incompatible with undistorted simulation.

So much for the development of simulation from Monte Carlo and the growth of high-speed computing machines. There is, however, a separate stream of simulation activity that now has partly merged with the Monte Carlo stream. Simulation is, of course, one of the most broadly defined words. It is used to describe things like Link trainers, radar simulation through use of models, various physical or electrical analogue devices, to things like those carried out by the RAND Simulation Laboratory--large-scale simulation of man-machine operations of an Air Defense Direction Center. Similarly, other people at RAND are now working on a simulation of the logistics system supplying and maintaining in operation several wings of, say, air defense fighters. Or, again, others have developed a Business Game which is a multistage, multiperson simula-

tion of a competitive business situation.* In all of the latter activities, people are engaged as players or participants as well as experimenters. Since there is an attempt to obtain realistic actions on their part, more stringent requirements are placed on the simulation to be like at least the most prominent parts of reality. This places considerable restriction upon the statistical designs that are permissable.

In the following two sections, there is a more concrete discussion of the statistical design problems in two cases: First, those cases in which simulation is of secondary, or of no interest. Here full range is available for the application of methods of increasing the accuracy of the estimates to be derived. Second, those cases in which simulation is important. Here the design problem is more confined, but there are several useful things that can be done to reduce the variances of the interesting and most important comparisons.

In the remainder of the paper some suggestions about new uses of techniques drawn from the statistical literature on the design of experiments are made. They seem to me the most applicable to the new area of Monte Carlo and Simulation. All of these suggestions are as yet untried.

II. Monte Carlo Design in Cases Where Simulation Interest is Secondary or Non-existent. This is not the place to write in detail of the main techniques that have been found useful in reducing the variance of estimates of Monte Carlo computations. First, the generalities for which there would be space are not too helpful. Second, there is now a substantial literature describing these techniques.**

The facts are that, even given the increasing speed of modern computing machines, it still is worth while to spend some time and effort in designing the calculations, and on the added coding time required in order to increase the accuracy of the estimates. Increases in accuracy on the order of several thousand-fold are not unusual. In other cases, however, after considerable hard work, factors of only 5 to 10 have been achieved. A characteristic of many of the problems where spectacular results are achieved is that there is some event of extremely low probability (such as the penetration of a shield by a neutron). An estimate of this probability to withing 10 per cent accuracy is desired. However, it is possible

---

*On the Construction of a Multistage, Multiperson Business Game, Bellman, Clark, Malcolm, Craft, and Ricciardi, JORSA, August 1957, pp. 469-503.

**One of the best places to read about these techniques is in Symposium on Monte Carlo Methods, H. A. Meyer, Ed., Wiley, 1956.

to force this event to have a reasonably high probability of occurrence (say, about .5 or .75) and to give it a weight (very small) each time it occurs. The average of all of these observations will have the same expected value as the original problem.

A problem for which the variance reduction so far achieved is small (on the order of five-fold) is that of estimating the average waiting time in a queue in which the arrivals are a Poisson process and the service times are distributed exponentially. This case can, of course, be solved analytically and has been studied as a Monte Carlo problem only for its methodological interest. None of the usual Monte Carlo tricks work. The only thing that has so far been discovered to give any substantial decrease in the variance of the estimated mean waiting time is the following sample scheme. Let $X_1, X_2, X_3, \ldots$ and $Y_1, Y_2, \ldots$ be sequences of inter-arrival times and service times. From these a sequence of waiting times can be generated, $w_1, w_2, w_3, \ldots$ and the average waiting time $\bar{w}$ computed. The sequences can be reversed and altered to be $\alpha Y_1, \alpha Y_2, \ldots$ and $(1/\alpha)X_1$, $(1/\alpha)X_2, \ldots$ so as to make the average inter-arrival time have the right relation to average service time. A second sequence of waiting times can be generated and $\bar{w}'$ computed. When $\bar{w}'$ is averaged with $\bar{w}$, each weighted equally, this new estimate is much better than either of the two original estimates. This is because the original estimates, $\bar{w}$ and $\bar{w}'$, are negatively correlated, as is easy to see. The interchange in the X and Y sequences is unusually easy in this case, but the idea has a general application.

The basic point is that design for the sake of variance reduction pays off and should be used more often than it is. However the application of the necessary techniques can seriously distort the simulation aspects of the calculations. For example, the use of what are called expected value methods completely destroys the simulation because part of the problem is done analytically. The use of importance sampling, splitting, and Russian Roulette, each of them devices for multiplying observations in interesting regions of low probability, distorts the simulation appropriate to the values of the parameters defining the problem to be studied. What would be typical observations are made into rare events and vice versa.

Many of the Monte Carlo variance reducing techniques are the same as techniques known to statisticians in the fields of design of experiments or sample surveys. Some of them seem to have been independently discovered by those interested in Monte Carlo. For example, importance sampling is equivalent to sampling according to size. The possibility of producing zero variance estimates through optimal choice of an importance sampling scheme seems to have been first noticed by the Monte Carlo people. Others of the techniques are definitely new--for example, Tukey's Conditional Monte Carlo and Kahn's hybrid-splitting. Some techniques have an altogether new importance in Monte Carlo, for example, the use of regression estimates and more generally the use of correlation to improve estimates. In ordinary experimentation one has to be satisfied with the amount of correlation nature has supplied. In Monte Carlo, the extent of correlation is to a large extent within our control through the careful use and re-use of the same sequence of random numbers. The latter leads to one of the important advantages of pseudo-random numbers over the real thing.

There are some ideas and techniques current in the area of design of experiments that have not yet been used, to my knowledge, in Monte Carlo calculations. Most interesting may be Box's ideas on how to look for maxima of response surfaces, a recurrent problem especially in operations analysis.*

Finally, a word about the special place of unbiased estimates in Monte Carlo. Everyone agrees that for an estimate to be unbiased is not in itself a recommendation. However, in Monte Carlo unbiased estimates play a very prominent role. Indeed, they are the only ones used. The reason is that:

1. The objective of the analysis is usually to estimate the mean value of some random variable, not the parameters of the distribution. We already know the values of the parameters, but not the function relating the mean value of the distribution to the parameter. Indeed we often cannot write down the distribution of the random variable, but we can sample from it. Since observations from the required distribution can be produced a natural estimate of its mean value is the sample average.

2. The easiest way to obtain a better estimate than the sample average is to hang on to the unbiased character of the estimate while using the added degree of freedom available in Monte Carlo computations to reduce the variance. For example, in the application of importance sampling, suppose an estimate of the mean value ξ of $g(x)$ is desired, where x is distributed with the probability density $f(x)$, i.e.,

$$\xi = \int_{-\infty}^{\infty} g(x)f(x)dx$$

the usual estimate would be

$$\overline{g(x)} = (1/N) \sum_{i=1}^{N} g(x_i) \quad .$$

A different estimate can be formed by sampling from another probability density $h(x)$ and weighting

$g(x_i)$ by $\dfrac{f(x_i)}{h(x_i)}$, i.e.,

$$\overline{g^*(x_i)} = (1/N) \sum_{i=1}^{N} g(x_i) \frac{f(x_i)}{h(x_i)} = (1/N) \sum_{i=1}^{N} g^*(x_i) \quad .$$

---

* G. E. P. Box, The Exploration and Exploitation of Response Surfaces: BIOMETRICS, March 1954.

The expected value of the new estimate is clearly ξ

$$\xi = \int_{-\infty}^{\infty} g(x) \frac{f(x)}{h(x)} h(x)dx$$

so long as h(x) satisfies a few mild restrictions. The variance of the new estimate can be made very small by proper choice of h(x), and indeed zero by choosing

$$h(x) = \frac{g(x)f(x)}{\xi}, \text{ if } g(x) \geq 0 .$$

Therefore unbiasedness is not valued for its own sake, consistency would do as well, but it seems the easiest thing to hold on to while manipulating the variance.

III. Monte Carlo Design in Cases Where Simulation is Important.

Nothing will be said here about the problems of physical stimulation, or of the special problems associated with simulation when human beings are used as players. These are problems for the physicist, engineer, and psychologist. In addition there is not much that can be said about the Monte Carlo design of these cases. The chances of doing a great deal through statistical design are compromised by the emphasis on simulation. Also, in the practical use of computing machines, the necessity in many simulation studies of taking out of the computer a large amount of data descriptive of the running of the model makes for additional design difficulties. This requirement, of course, limits the effectiveness with which high-speed machines can be used.

Nonetheless some variance reduction can be achieved through proper design without distorting the simulation. Suppose the problem, typical in operations analysis, is that of deciding whether system A is better than system B, and by how much. System B may be the system in current operation. Suppose also that it is desired that the parameter values believed to be those of the real life situation be used so as to produce a real life simulation. Without in any way distorting the simulation of A or B, the use of correlated random variables throughout the simulation of the working of each system will improve the comparison of A and B, and give an improved estimate of the potential gain in adopting A. The amount of reduction in the variance of the estimate of the difference between A and B will depend on the problem. In typical problems, a small amount of effort put on arranging the correlation gives results of about 25 to 50 fold reduction in variance.

Correlation techniques have a wide range of application. The reversing and re-using of sequences of random numbers as described in connection with the queueing problem discussed earlier is a correlation technique that would sometimes be useful in a different type of problem. In other cases it may be advantageous to introduce artifically a second model or system; often a simpler model of the situation under study and for which it is

possible to analytically calculate its mean value. The essential feature is that the mean value of the second, artificial problem be known. Then correlated sampling of both problems can be exploited to produce by means of a regression estimate an improved estimate of the mean value associated with the problem in which we are really interested.

Another technique that seems to be useful in simulation problems is the mild or occasional use of importance sampling in order to force interesting events to happen more frequently than they otherwise would. For example, although a good inventory policy is one designed so that certain costly events happen very infrequently, a Monte Carlo Simulation intended to test the performance of the policy might be designed to force these catastrophic events to occur in order to see what happens, how costly they really are, what kind of recovery the system makes, etc.

Currently the RAND Logistics Laboratory is using both these techniques on their problems.

AN INSTRUMENT FOR THE DETERMINATION OF IMPACT SENSITIVITY OF MATERIALS IN CONTACT WITH LIQUID OXYGEN

William R. Lucas and Wilbur A. Riehl
Army Ballistic Missile Agency

ABSTRACT. An apparatus for determining the impact sensitivity of materials in liquid oxygen is described. The impact tester provides flexibility of testing conditions and reproducible results. Test results are presented illustrating variables which must be controlled in impact testing.

INTRODUCTION. Liquid oxygen is one of the most important oxidizers in liquid rocket propellant systems. Pure liquid oxygen (LOX) is stable and not subject to detonation by mechanical shock, but mixtures of LOX with most organic materials and certain inorganic materials including aluminum, magnesium, lead, and iron oxide will explode under conditions of impact. Spontaneous combustion does not occur generally in the liquid phase, because of the extremely low temperatures involved. However, under mechanical shock, detonation will occur when LOX contacts many elastomers and gasket materials, sealants, lubricants, threading compounds, and hydro carbon residues on components. Aluminum valves and lines transporting LOX have been observed to explode and burn like wicks, the only explanation being greasy fingerprints left during assembly. During testing of missiles, there have been explosions resulting from contact between LOX and valve lubricants or fitting sealants which jeopardized the whole system. The Ordnance Safety Manual precludes the use of organic material with LOX. Of course, this is impractical in missile systems. Therefore, in order to classify materials as to degree of hazard and to provide a means of qualifying materials for use with LOX, an impact sensitivity test apparatus was developed. Furthermore, it is not sufficient to certify a given product for use with LOX, but it is necessary to test each batch of material to be used with LOX, especially heterogeneous material from which some gaskets are made.

Impact testers are in common usage for testing explosives, and at least two other testers have been used for qualifying materials for use with LOX. Impact testing is empirical at best, and it became obvious to these authors early in their experience of testing materials with crude testers for use with LOX that it was imperative to design an instrument for maximum reproducibility. Some of the early instruments were more variable than the materials being tested. Impact levels for acceptability can be set arbitrarily by calibrating the instrument against a few materials known to be reasonably safe with LOX. However, an instrument must give reproducible results day after day, and two or more instruments of the same design must give reproducible results. With these considerations in mind, the instrument described herein was developed.

DESCRIPTION OF INSTRUMENT AND TEST PROCEDURE. The instrument consists of a plummet guided in its vertical fall by two sets of ball bearings, one set at each end of the plummet and arranged at the vertices of equilateral triangles, which roll freely in tracks milled in 1" x 1" x 72" stainless steel bars, see figure 1. These tracks are bolted rigidly to unistrut members and accurately aligned with shims, providing for even contact with the ball bearings at all points along their length. The unistrut supports

are securely anchored to the top and base plates as indicated in figure 1.* The base plate is mounted with leveling screws, to a steel frame of table height, which in turn, is secured to the concrete floor.

The plummet is held at the desired height by an electromagnet, supplemented by a spring-loaded safety catch in positive action at all times except when current is delivered to the solenoid activating the release of the safety catch. Figure 2 shows the electromagnet and the suspended plummet. The electromagnet may be positioned at any height from 0 to 127 cm and the total effective drop distance read directly from calibrations on the electromagnet support shaft as shown in figure 2.

Located on the electromagnet support and to the left of the locking handle is the height indicator. This must be reset to read zero cm with the plummet resting on a striker pin each time a change is made in plummets and / or sample cup holder assemblies. Then the drop distance of the plummet can be read directly at any electromagnet setting.

Two plummets are used, one weighing 9 Kg and the other weighing 3.4 Kg. These provide operating ranges of 0 to approximately 11 KgM and 0 to approximately 4 KgM respectively. The range is chosen which best suits the sensitivity of the material under investigation.

The plummet is released by simultaneously depressing two buttons on the control panel, one button releasing the safety catch and the second reversing the field of the electromagnet so that it will release as the polarity nears zero. The plummet delivers the impact to a 2" x 1/2" diameter stainless steel striker pin resting on the sample and immersed in LOX. Figure 3 shows the plummet in the impact position. As indicated in figure 4, the striker pin is held in position by a stainless steel collar, sliding on two guide pins mounted in the base plate. The stainless steel sample cup holder is held to the base plate by spring clamps. This cup holder is interchangeable with a liquid nitrogen moat, serving the same function, but permitting cooling of the sample below the boiling point of LOX. The moat consists of a stainless steel box to which liquid nitrogen may be added manually by pouring or by attaching the moat directly to a 50-liter Dewar flask and transferring by pressure as required. The moat was designed primarily to conserve LOX. However, the safety hazard introduced by its use is not justified by the advantage gained. Thus, the moat is very seldom used.

Figure 5 shows details and orientation of striker pin, sample cup, and sample. A clean striker pin is used for each test, thus several clean pins are necessary for a series of tests. The sample cups are stamped from commercially pure aluminum and are used for only one test, then discarded.

The control panel, shown in figure 1, is separated from the instrument by a barricade containing an observation window arranged so that the operator has a view of the sample cup. However, the instrument should be as near the barricade as possible so the minimum time elapses between LOX topping

* Figures appear at the end of the articles.

and impact. The test cell which is lined with acoustical tile is darkened during a test to facilitate observation of sparks or a flash. The Universal Counter shown in figure 1 is used to measure the time interval as the plummet drops between fixed points and thus to evaluate friction loss. Figure 6 is a comparison of theoretical acceleration of the plummet at Redstone Arsenal and the measured acceleration on the properly aligned instrument. As shown in this figure, frictional loss is slight and can be considered negligible. This test should be repeated periodically as a check of the tester.

The nature of the sample determines the manner in which it is prepared for testing. Solids are cut into discs smaller in diameter than the inside dimension of sample cup and greater in diameter than the striker pin. Samples are cut without cutting oil or other coolant if possible and are brushed free of dust or gragments prior to testing. Single pieces of gaskets are tested so as to parallel the normal gasket environment. Oils, greases, and sealing compounds are tested as smears in the bottom of the test cup. The effect of the thickness of these smears will be discussed later in the report.

It is imperative that the instrument, its accessories and the room in which tests are done, especially the ceiling, be clean. Where lubricants, oils or other materials which may splash upon impact are tested, the tester must be cleaned between tests on different materials by scrubbing with steel wool and rinsing with a pure, chlorinated hydrocarbon solvent. This includes the base plate, guide tracks and plummet nose. The base plate may be covered with aluminum foil during test to simplify cleaning. A striker pin must not be used more than one time without cleaning. The pin is cleaned by vapor degreasing and alkaline cleaner soak, followed by thorough rinsing. The face of the striker must be free of pits and scratches. A test cup is used only once and discarded. Prior to its use, it is cleaned by scrubbing in detergent, water rinsing, and finally a rinse with pure, distilled chlorinated hydrocarbon solvent. After cleaning, the sample cup and striker pin are never handled by hand, and forceps, tongs, sample trays, etc., are cleaned in the same manner as the striker pins and sample cups. Striker pins are precooled to LOX temperature by suspending on a clean, stainless steel wire in a Dewar flask of LOX. Samples and cups are precooled by immersion in a separate container of LOX for each sample. The cup is full of LOX before transferring to the tester and the cup is topped with LOX just prior to impact.

Samples are tested at a given impact level and this level reduced until in twenty consecutive tests there is no evidence of sensitivity. Sensitivity is indicated by an audible report or a flash visible in the darkened room. The level at which there is no evidence of sensitivity in twenty consecutive tests is called the level of LOX impact insensitivity.

TEST RESULTS. The tester described represents the third refinement of a rather crude beginning in impact testing of materials for contact with LOX. The current tester has been used for approximately three years and has given consistent results on different types of materials such as lubricants, oils, greases, solvent residues, gaskets, etc. Recently, the instrument has been used for determining impact sensitivity of materials in concentrated hydrogen peroxide.

Inasmuch as more work has been done recently on lubricants and sealants than on any other material, test results from this type of material will be discussed briefly in order to identify testing variables which must be controlled.

In the early days of testing, insensitivity level of material tested was established on the basis of ten consecutive tests. However, there have been cases where there were no detonations in the first ten trials but one or more detonations in the second ten trials. In a consideration of 1760 individual tests, it was shown that with duplicate tests, one can expect the number of detonations in the first set of ten to agree within $\pm$ 1.5 of the number of detonations in the second set of ten trials, two out of three times. In several series of twenty duplicate tests the number of detonations in the first twenty trials agreed within $\pm$ 1.5 detonations of the number in the second twenty, two out of three times. Obviously, with the greater number of tests, the reproducibility of results is increased, however it seems impractical to make more than twenty consecutive test at one impact level on a given material.

An acceptable impact level cannot be established which would be applicable to all materials. Where the anticipated impact in use is known, this level with a safety factor is established as the acceptable test level. In most cases, the environment cannot be evaluated quantitatively, and it is necessary to select as the standard the impact insensitivity level of a material which has been used successfully. For example, Oxyseal, a commercial sealant consisting primarily of a mixture of graphite and a chlorinated biphenyl, has been used in rocket engines for several years on valves, pipe fittings, etc., in LOX service. Since the experience of several agencies using this material has been favorable and no explosion has been traceable to it, this material constitutes a standard against which to check other sealants.

A series of tests was made in order to show the effect of sample thickness on test results. Some of the results are shown in Table I in which it is seen that reducing the specimen thickness with all other factors being constant increased sensitivity by sixteen times. In another test of a proprietary sealant at 7 KgM force, there was one detonation in twenty trials using a sample thickness of .050 inches and under the same conditions except with a sample thickness of .004 inches, there were nineteen detonations in twenty trials. The thickness of the sample can be checked easily by weighing the sample cup before and after addition of the sample and then calculating the depth, knowing the density of the material being evaluated. Since the purpose of the impact test is to qualify a material for use with LOX, it seems desirable to select a test condition which is most sensitive to variations in materials. In this case, the thicker sample seems to provide the better results. Whereas with the thin sample the spread between an acceptable material, Oxyseal, and an unacceptable material, Proprietary Sealant No. 1, is only 2 KgM, the spread using a thick sample is approximately 8 KgM.

The effect of the plummet mass on test results was observed. In one case, a plummet weighing 9.04 Kg was used to provide a given impact force and in the other case a plummet weighing 3.4 Kg was used to provide the same impact force. Results of this series of tests are presented in Table II. As can be seen from this table, the lighter plummet indicated a higher order of sensitivity as expected.

In the ABMA tester, the diameter of the striker pin is one-half inch. The effect of varying this diameter when testing gasket materials is shown in Table III It is shown that the insensitivity level essentially varies directly with area of striker pin face. There is not a direct correlation between impact insensitivity of lubricants and area of impact, however the importance of using a constant impact area is established.

The effect of cleaning techniques for sample cup and striker were investigated and results are shown in Table IV. Method A of Table IV is the standard practice Thus, it is evident from these results that much care must be given to cleaning and handling the sample. By placing a few particles of sand, alundum or carborundum as fine as 320 mesh in a sample cup with LOX, the aluminum cup was made to react explosively under impact in 50% or more of the trials. This emphasizes further the importance of cleaning.

In order to show the manner in which sensitivity decreases toward a level of insensitivity, data on a series of representative samples are shown in figures 7, 8, 9, and 10. From these results, which show that in some cases a wide range exists between 100% and 0% activity, it appears dangerous to establish an acceptability level of impact which allows any detonations in twenty trials. Therefore, acceptance levels are established at impact forces at which no detonations are experienced.

A variety of materials applicable to the oxidizer systems of missiles has been tested with the instrument described herein. On the one hand are the comparatively insensitive materials such as Teflon, Kel-F, fluorolube, polyethylene, and on the other hand are the very sensitive materials such as leather, epoxy resins, hydraulic oils, some commercial sealants, and hydrocarbon residues from cleaning solvents. The responses of the various sensitive materials to impact vary widely. In some cases, there is a loud report from the approximately 0.5 gm sample, accompanied by a flash, and often the surface of the aluminum cup is melted. In other cases, there is only a flash. There is only evidence of slight charring in some of the less sensitive materials.

CONCLUSIONS. In conclusion, a test apparatus was developed for determining the impact sensitivity of materials in LOX. This apparatus has the following advantages over other known instruments designed for the same purpose:

1. Provides reproducible results.
2. The low friction loss of the falling plummet approaches, in effect, a free falling body.
3. The striker pin is immersed in LOX and in contact with the sample at time of impact. This reduces splash encountered when the nose of the plummet is the striker.
4. Utilizes cheap, expendable sample cups.
5. By using expendable sample cups and a clean striker pin for each test, both impact surfaces are fresh for each test.
6. Provides flexibility of testing conditions to accommodate widely varying material.
7. Provides adequate safety features.

The chief feature desired in an instrument of this type is reproducibility of test results. It is believed that other instruments built according to the plans for the ABMA instrument and used according to the standard operating procedure will provide the same results. Although impact testing is not the only consideration for selecting materials to be used with LOX, it is certainly one of the most important.

The ABMA impact tester has been recommended by the Associate Contractors of Ramo-Wooldridge Corporation, the Army Ballistic Missile Agency and the Wright Air Development Command as the standard test equipment for determining the impact sensitivity of lubricants and sealants used in a LOX environment. A test procedure involving this apparatus constitutes a part of the tentative specification, "Lubricant, Antiseize and Sealing, Liquid Oxygen Systems," which is patterned after the comparable Air Force and Navy BuAer-approved Military Specification MIL-T-5542B(ASG), "Thread Compounds, Antiseize and Sealing, Oxygen Systems," applicable to gaseous oxygen systems.

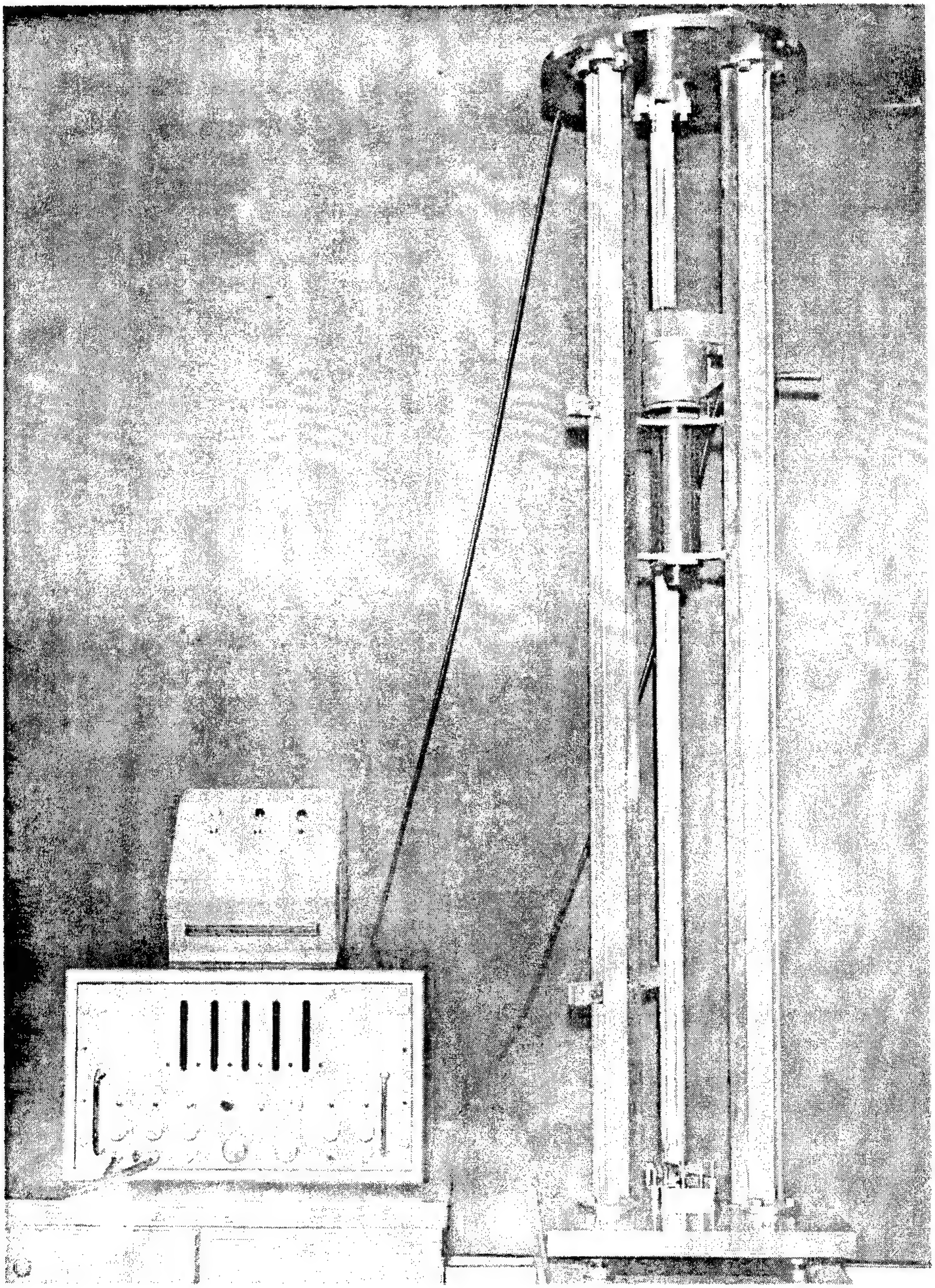


Fig. 1 Impact Tester

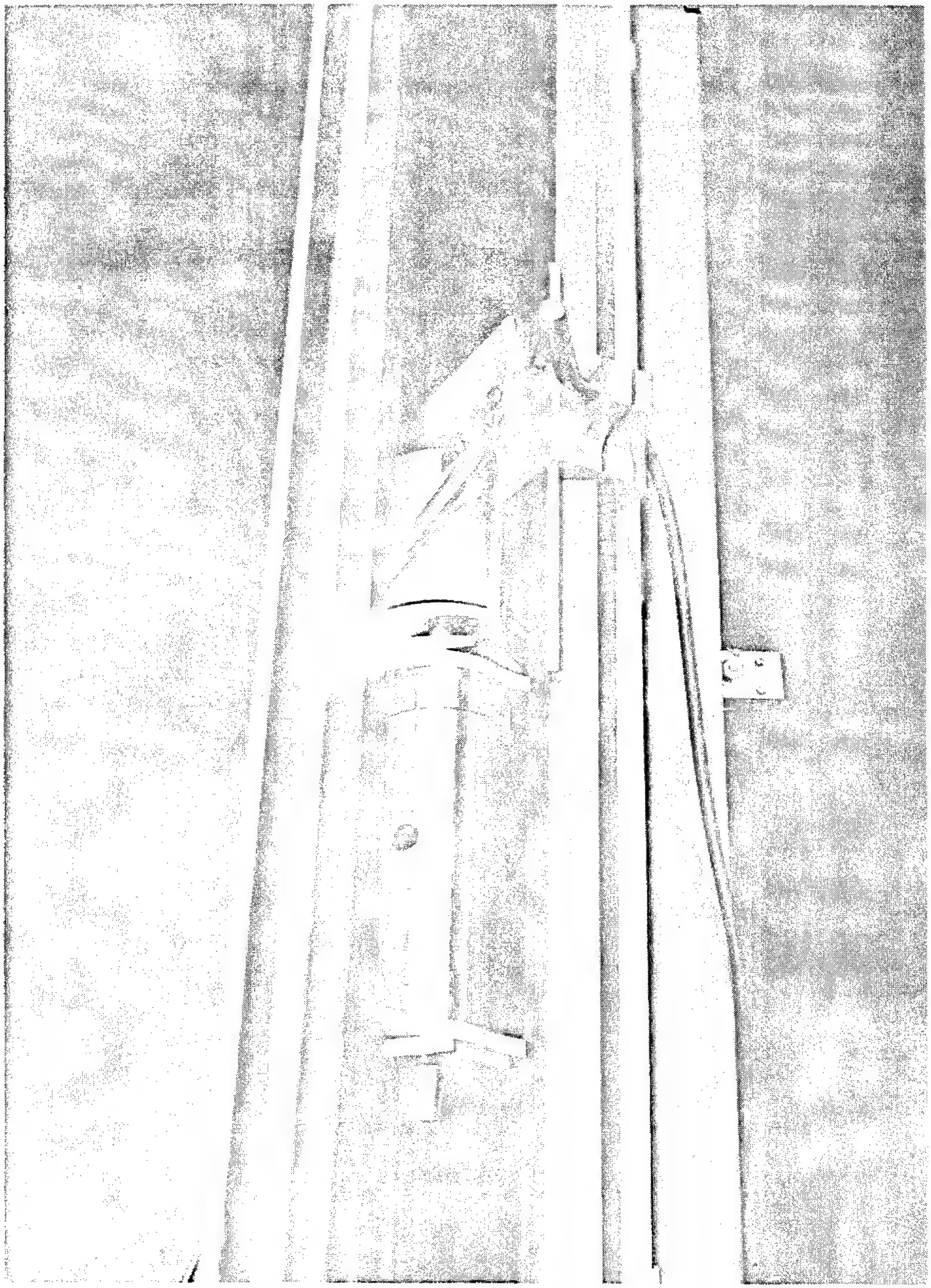


Fig. 2 Impact Tester Plummet Release Mechanism

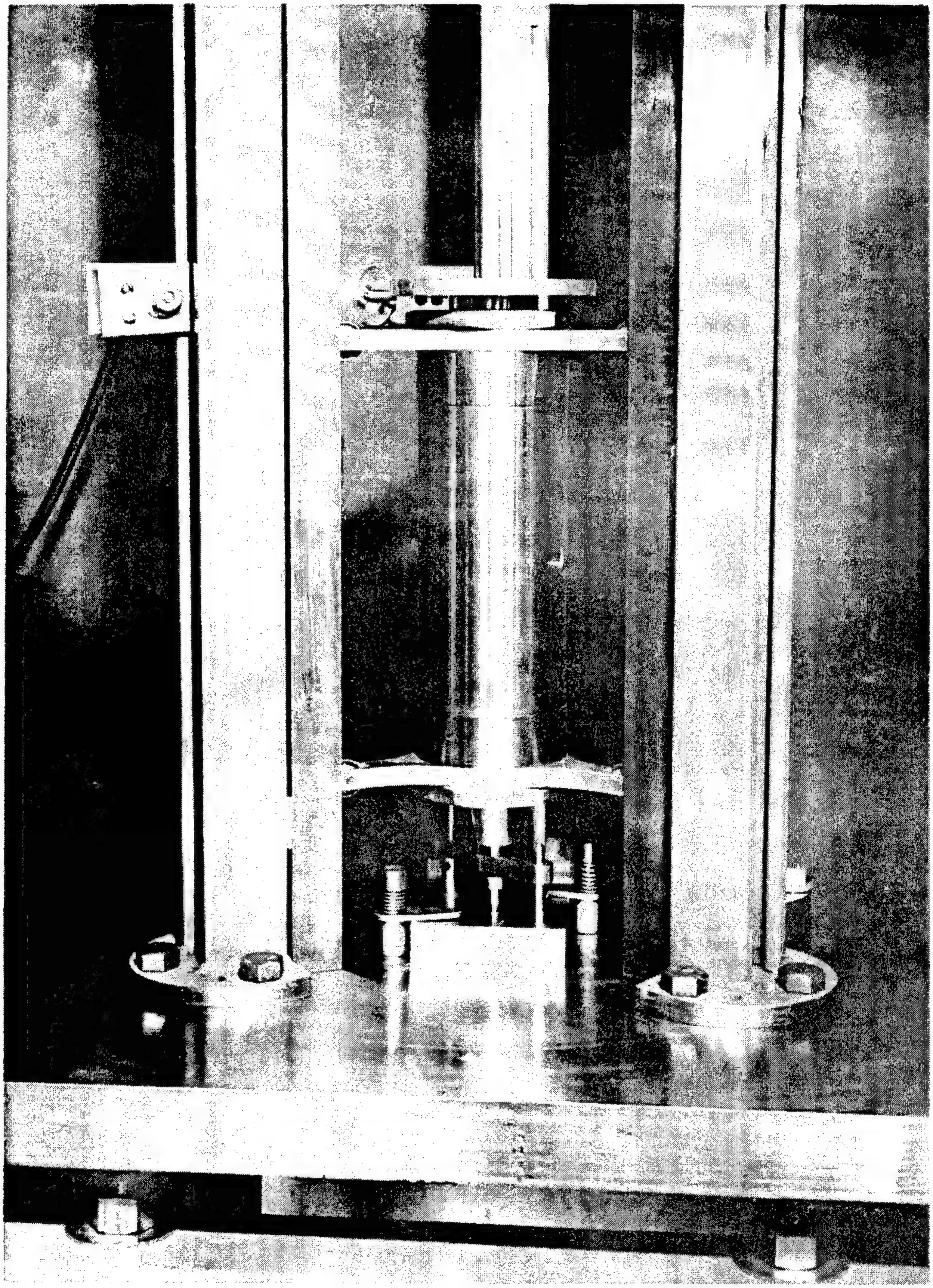


Fig. 3 Plummet in Impact Position

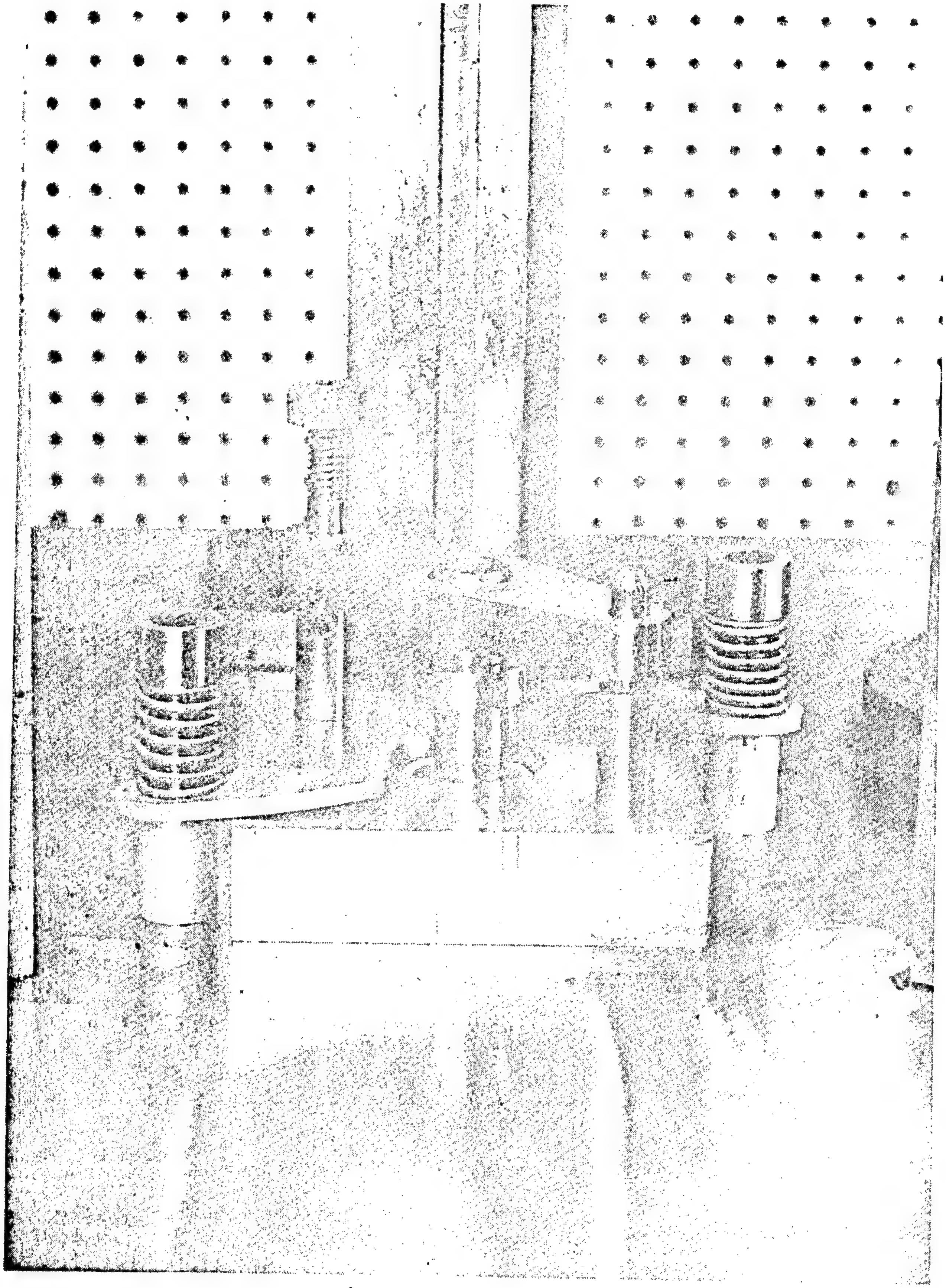


Fig. 4 Striker Pin and Sample Cup Assembly

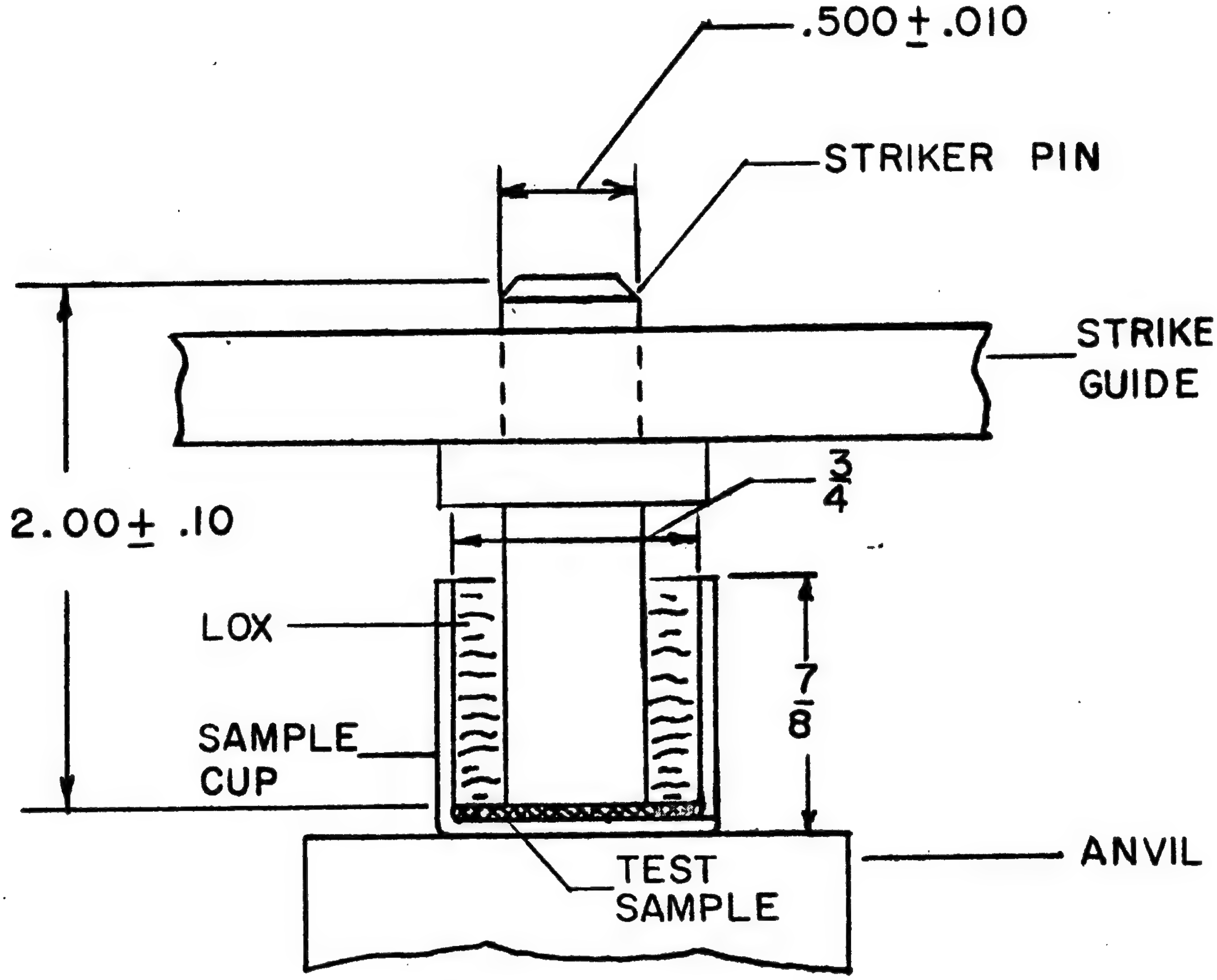


FIG. 5 DETAILS OF STRIKER, SAMPLE CUP, AND SAMPLE

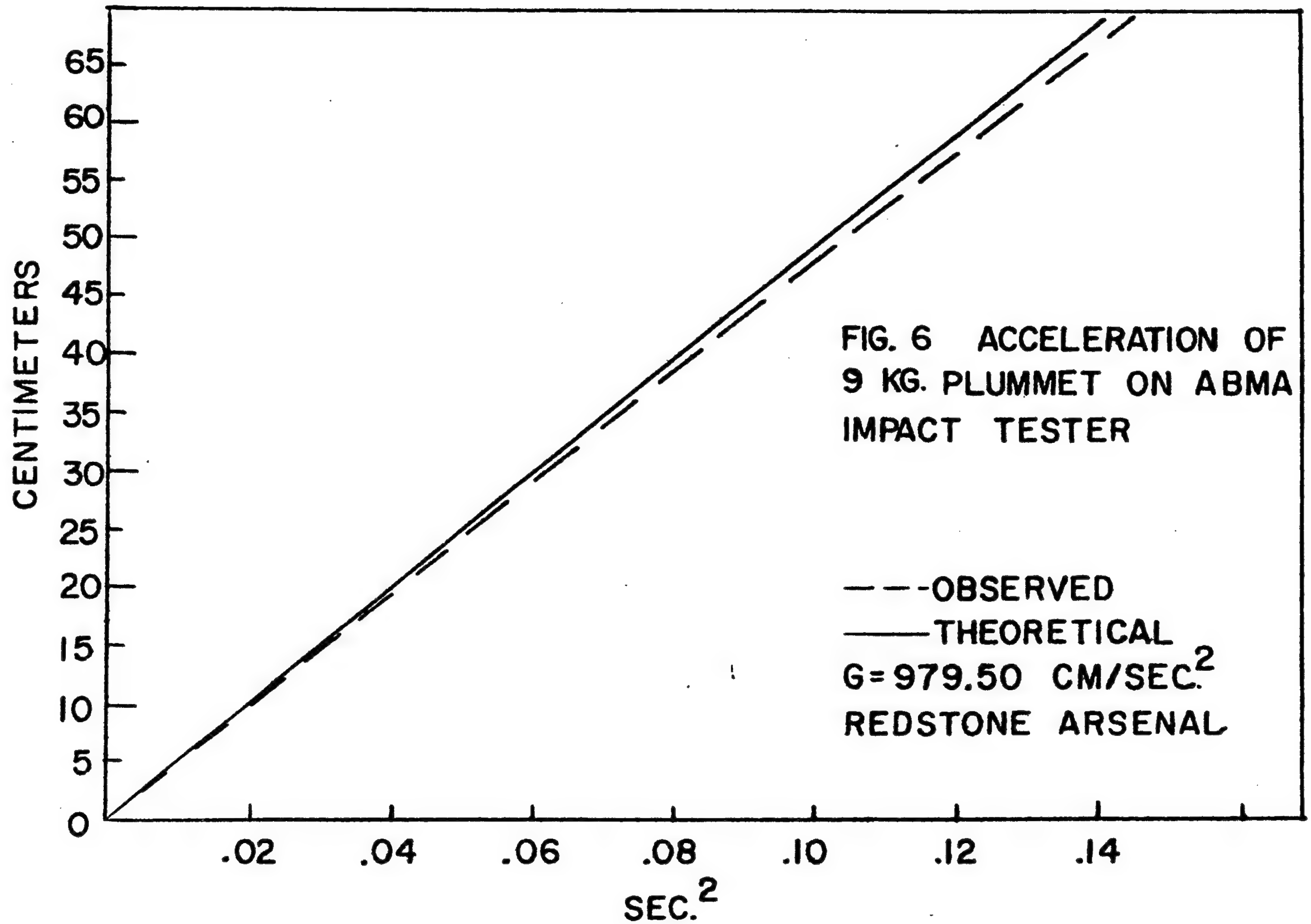


FIG. 6 ACCELERATION OF 9 KG. PLUMMET ON ABMA IMPACT TESTER

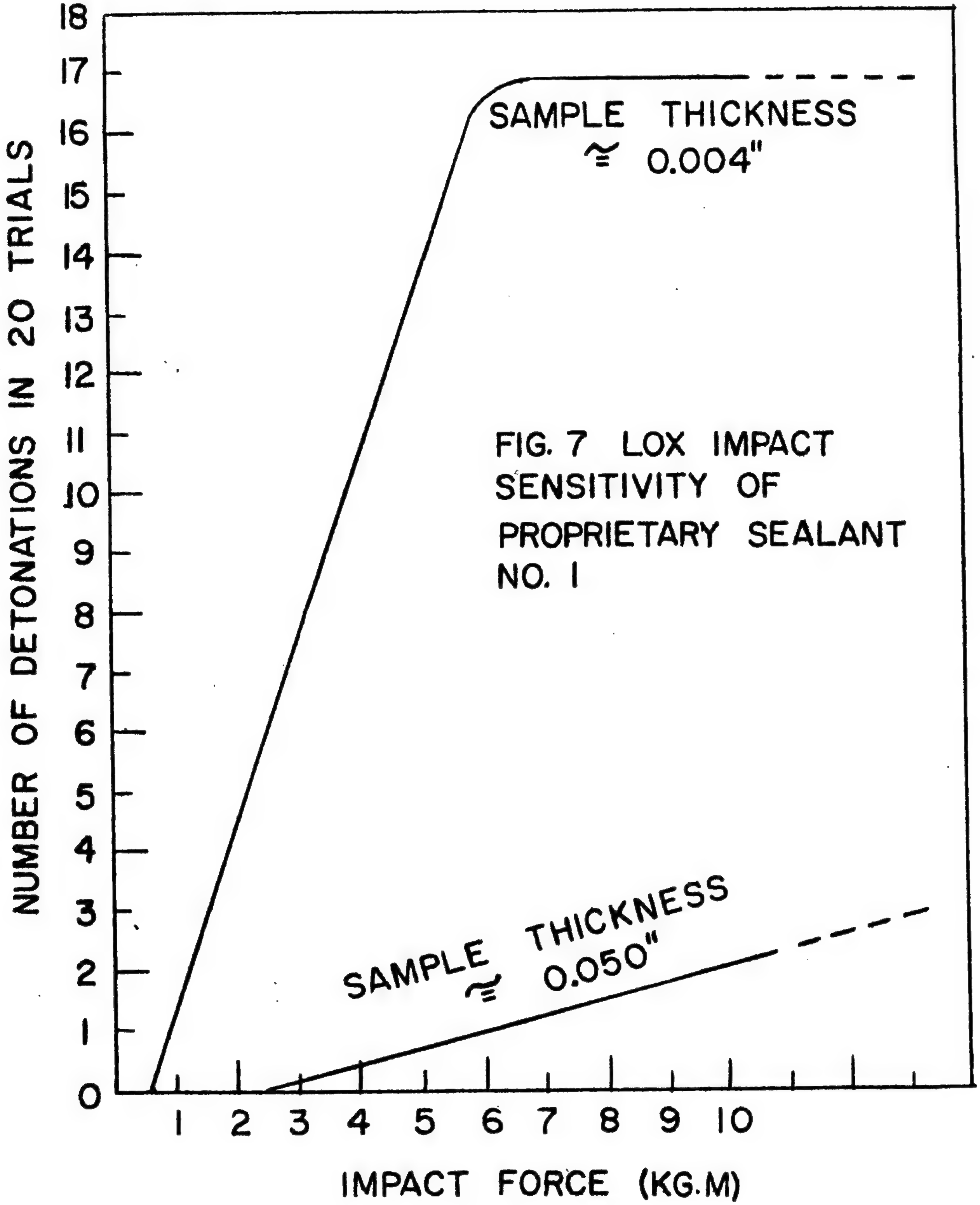


FIG. 7 LOX IMPACT SENSITIVITY OF PROPRIETARY SEALANT NO. 1

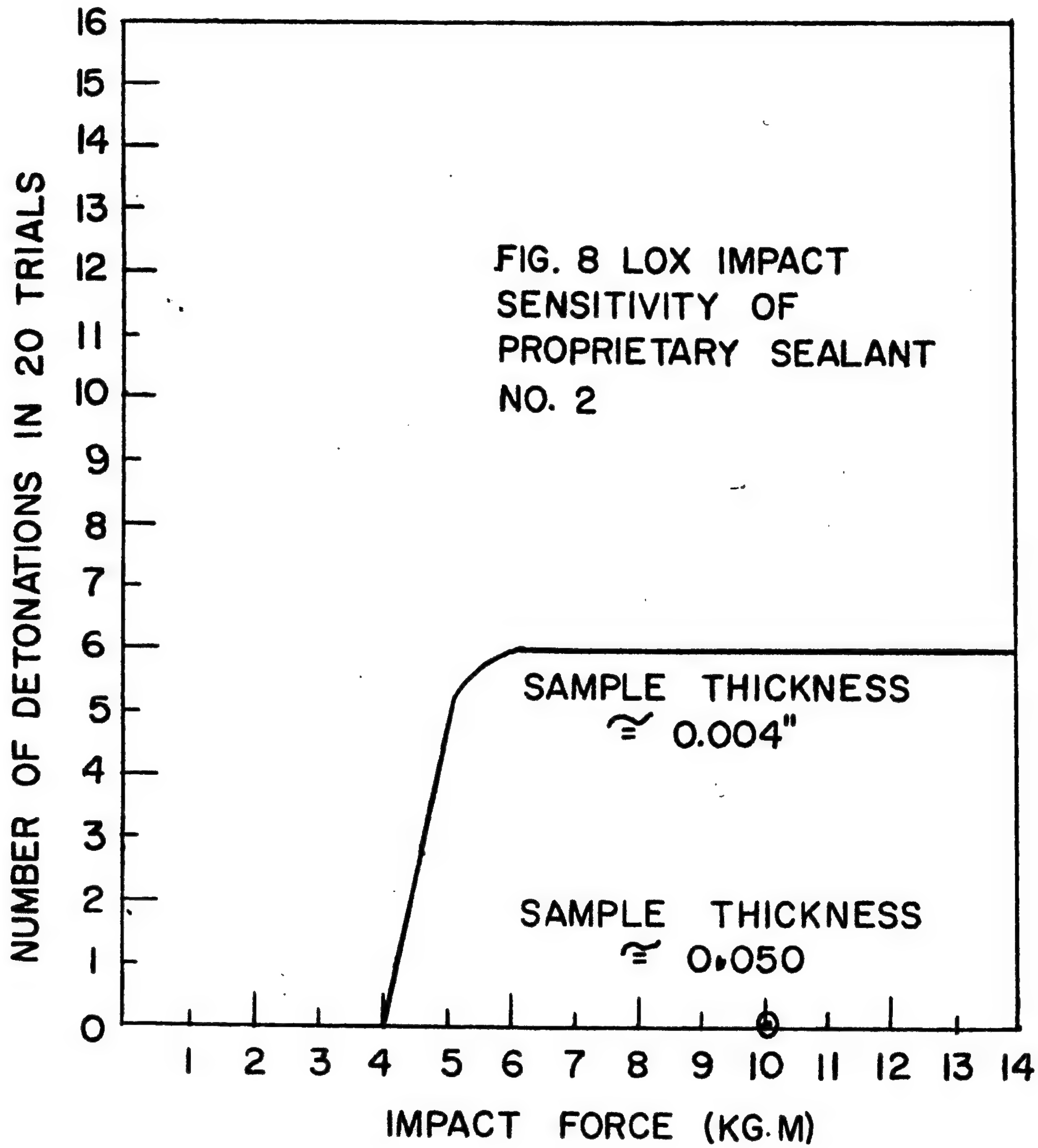
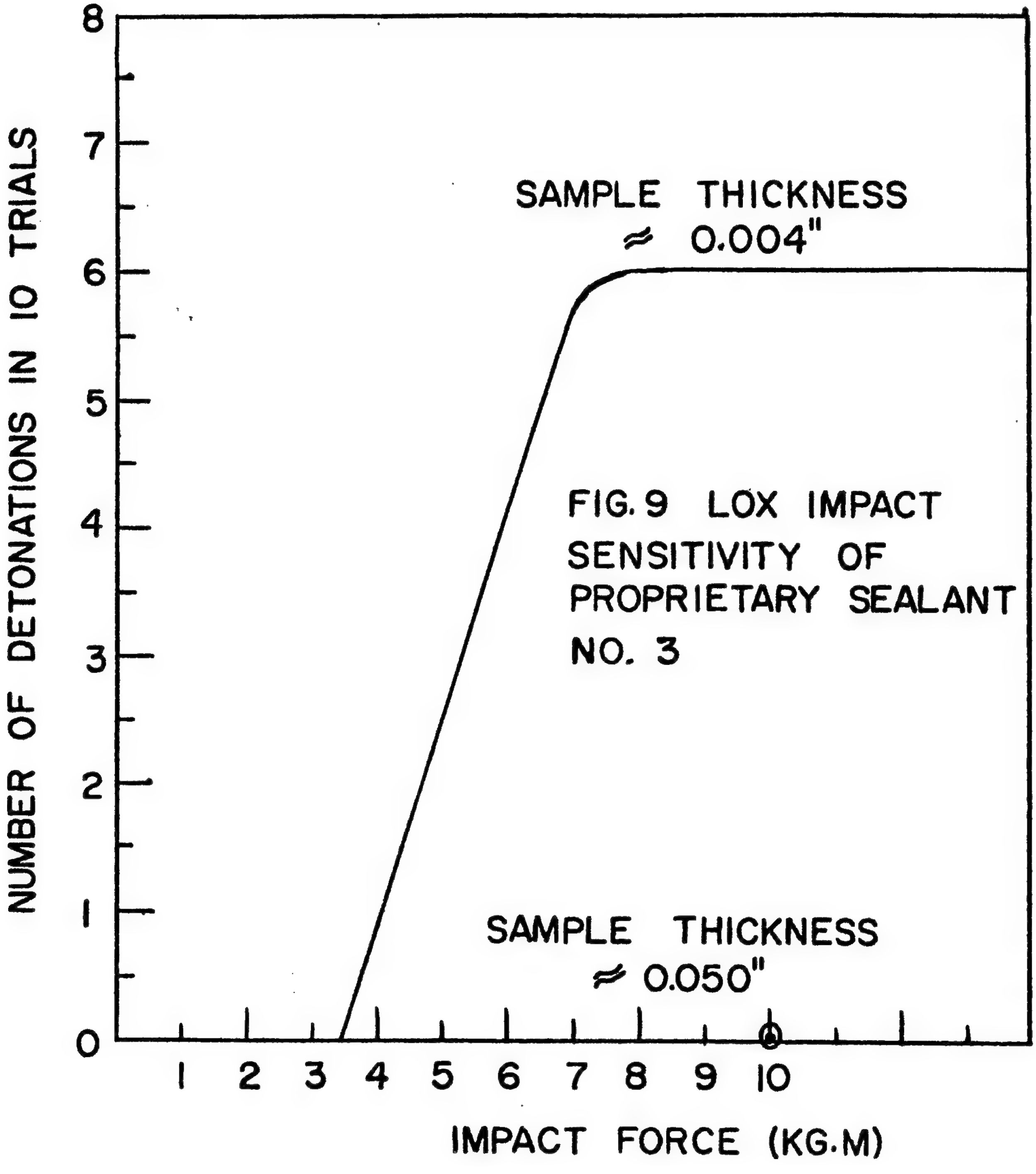


FIG. 8 LOX IMPACT SENSITIVITY OF PROPRIETARY SEALANT NO. 2
NUMBER OF DETONATIONS IN 20 TRIALS
0
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
SAMPLE THICKNESS ≅ 0.004"
SAMPLE THICKNESS ≅ 0.050
1
2
3
4
5
6
7
8
9
10
11
12
13
14
IMPACT FORCE (KG.M)


NUMBER OF DETONATIONS IN 10 TRIALS
8
7
6
5
4
3
2
1
0
SAMPLE THICKNESS
≈ 0.004"
FIG. 9 LOX IMPACT
SENSITIVITY OF
PROPRIETARY SEALANT
NO. 3
SAMPLE THICKNESS
≈ 0.050"
1
2
3
4
5
6
7
8
9
10
IMPACT FORCE (KG.M)

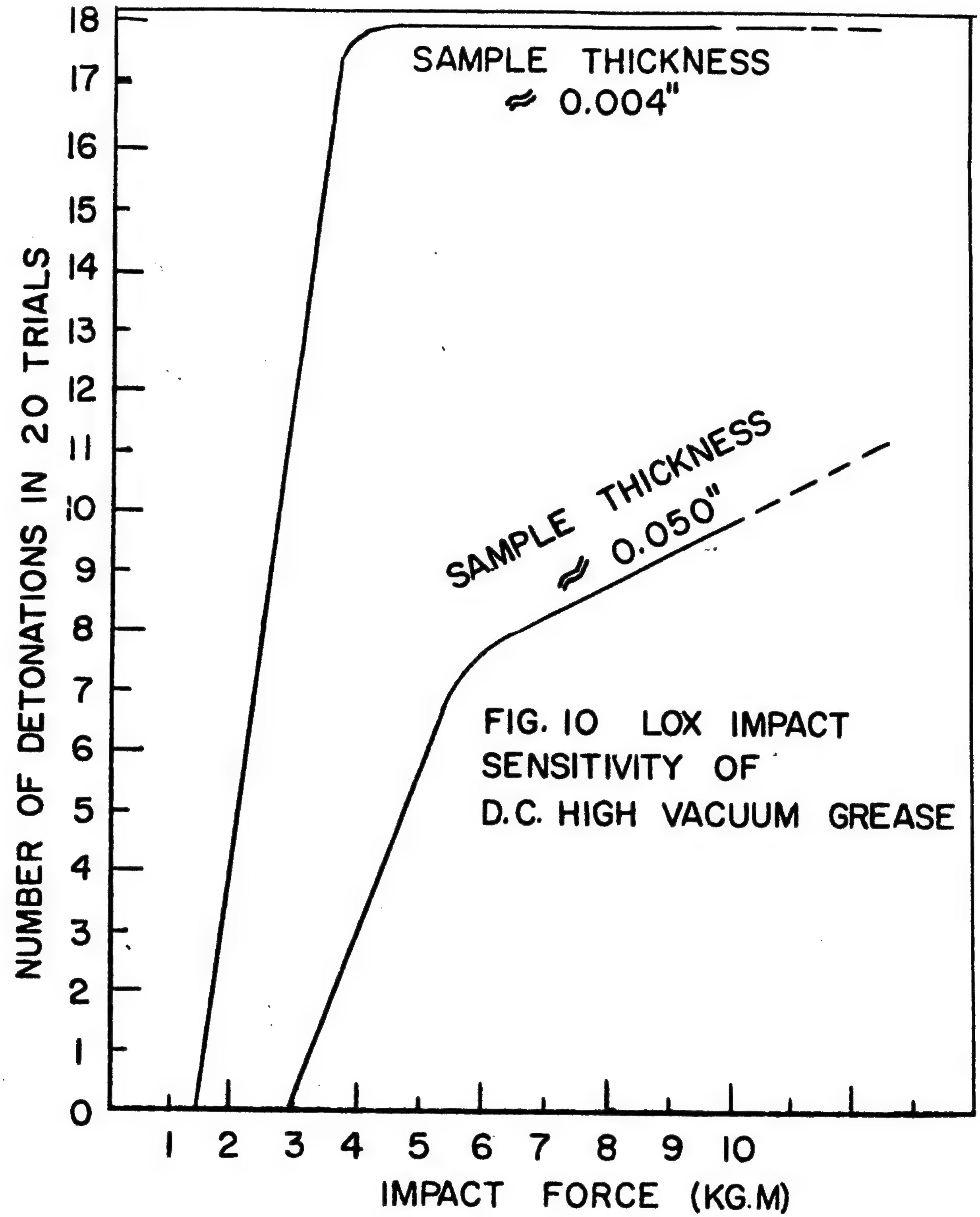


FIG. 10 LOX IMPACT SENSITIVITY OF D.C. HIGH VACUUM GREASE

TABLE I

EFFECT OF SAMPLE THICKNESS OF A PROPRIETARY SEALANT ON LOX IMPACT SENSITIVITY

	SAMPLE THICKNESS $\approx$ .050"	SAMPLE THICKNESS $\approx$ .004"	
IMPACT FORCE	NR DETONATIONS	NR DETONATIONS	INCREASE IN ACTIVITY
4 KG.M (9.04 KG PLUMMET)	1/40[x]	8/20	16X
4 KG.M (3.4 KG PLUMMET)	2/40	16/20	16X
10 KG.M	4/40	29/40	7.25X

[x] INDICATES ONE DETONATION IN FORTY TRIALS

TABLE II

EFFECT OF PLUMMET MASS ON LOX IMPACT TEST RESULTS

MATERIAL	IMPACT FORCE KG.M	DETONATIONS 9.0 KG. PLUMMET	DETONATIONS 3.4 KG. PLUMMET	INCREASE IN ACTIVITY
PROPRIETARY SEALANT NO. 1 .050" THICK	4	1/40[x]	2/40	2X
PROPRIETARY SEALANT NO. 1 .004" THICK	4	8/20	16/20	2X
PROPRIETARY SEALANT NO. 1 .050" THICK	3.5	1/20	5/10	10X
PROPRIETARY SEALANT NO. 1 .050" THICK	2.5	2/20	4/10	4X
PROPRIETARY SEALANT NO. 2 .004" THICK	4.0	0/20	1/20	

[x] INDICATES 1 DETONATION IN 40 TESTS

TABLE III

EFFECT OF DIAMETER OF STRIKER FACE
ON LOX IMPACT TEST RESULTS USING THE ABMA TESTER

GASKET MATERIAL	INSENSITIVITY LEVEL KG.M 1/2" DIAMETER	INSENSITIVITY LEVEL KG.M 1/4" DIAMETER	INCREASE IN INSENSITIVITY LEVEL
ALLPAX 5364 A	5	9	1.8X
ALLPAX 5408 A	5	10	2X
ALLPAX 5408B	5	9-1/2	1.9X
ALLPAX 5409B	5	9-1/2	1.9X
ALLPAX 7700	4-1/2	7	1.6X
ALLPAX 9662	4-1/2	9-1/2	2.1X
ALLPAX 9670	5	10	2X
JOHNS-MANVILLE # 76	4.5	9	2X
JOHNS-MANVILLE # 60	3	8	2.7X

TABLE IV

EFFECT OF CLEANING TECHNIQUES ON LOX IMPACT SENSITIVITY[x]

CLEANING TECHNIQUES

CUP	STRIKER PIN	DETONATIONS IN 40 TESTS
A. DETERGENT[xx], H_2O RINSE, CCL_4 RINSE	VAPOR DEGREASE + ALKALINE CLEAN	0
B. VAPOR DEGREASE + ALKALINE CLEAN	VAPOR DEGREASE + ALKALINE CLEAN	2
C. DETERGENT, H_2O RINSE, CCL_4 RINSE	VAPOR DEGREASE + CCL_4 SCRUB	4
D. DETERGENT, H_2O RINSE, CCL_4 RINSE	STEEL WOOL + VYTHENE	6

[x] IMPACT OF 10 KGM THROUGH A STRIKER OF 1/2" DIAMETER FACE IN SAMPLE CUP FILLED WITH LOX

[xx] SORAPON SF

EXPERIMENTAL INVESTIGATION OF THE RESPONSES OF A LIQUID IN AN OSCILLATING CONTAINER

Werner R. Eulitz and Herman Beduerftig
Army Ballistic Missile Agency

Part 1*
ANALYTICAL CONSIDERATIONS

1. INTRODUCTION. The effect of a coupled oscillating system, where a liquid is forced to oscillate, is of great importance in a liquid propellant missile, principally because of its influence of flight stability. In this connection, the term "SLOSHING" is commonly used. It is associated with such daily occurrences as the splashing of a cup of coffee or of a full pail of water and is considered to be the point of resonance where the forced oscillation is equal to the natural frequency of the liquid. There are several publications concerning the subject of sloshing in which the approach is primarily theoretical, but it has not been possible to draw any direct conclusions for a practical method by which to damp sloshing.

It is generally acknowledged that an oscillating liquid may be considered as partly a free oscillating mass and partly a rigid mass, and this concept can be used in developing a practical method to counteract sloshing. From theoretical considerations of this division of the liquid mass, I derived the depth of the free oscillating mass to be about one-fourth of the tank diameter, when the tank has been filled to a height greater than the tank diameter. Based on this, it should be possible to damp the sloshing and stabilize the liquid by surpressing the liquid motion in this free oscillating mass.

One possibility for damping the free oscillating mass is by changing the natural frequency of the liquid. (Slide 1)** The formula for the natural frequency of a liquid is

$$\omega_n^2 = \frac{2g\Sigma_n}{d} \quad \tan h \; \frac{2\Sigma h}{d}$$

Where ω_n = $2\pi f$ natural frequency
g = acceleration
Σ_n = Bessel function
d = tank diameter
h = height of liquid

This slide is limited to the first mode. If the tank ratio of h/d is not too small, the natural frequency is dependent on the rank diameter only, since the hyperbolic tangent approaches 1. The value for the hyperbolic tangent is practically 1 and natural frequency is constant at all points where the tank ratio h/d is greater than or equal to 0.5.

The lower curve shows the relationship between the natural frequency and the tank diameter. As the tank diameter increases, the natural frequency decreases, and vice versa. By dividing the dangerous oscillating mass into many smaller parts, (Slide 2) the natural frequency of a single part increases to the point where the natural frequency of the original tank is no longer effective. This led to the use of the so-called "egg crate baffle," which was tested by another activity and found successful

---

*Authors of Parts I and II are respectively W. R. Eulitz and H. Beduerftig.

**See page 53.

in its damping effect. We discarded consideration of this device, however, because of the excessive weight that would be added to the missile.

In order to find a satisfactory damping device for use in missile propellant tanks, we have performed numerous experimental investigations, and I now propose to present a review of some of our accomplishments and those results which I believe to be of interest to you.

2. TEST EQUIPMENT. All tests were made in Plexiglass tanks of various sizes, using water as the liquid, (Slide 3). The tank shown is 17.5 inches in diameter and has a natural frequency of 1.43 cps. It is suspended like a pendulum, with an amplitude of 0.5", that is, with a 1.0" stroke. The oscillation is produced by a stepless, variable speed gear drive; an eccentric; and a shaft. With this drive, the frequency range is from 0.6 to 2.5 cps, the natural frequency of the tank being approximately midway.

The movement of the liquid at different frequencies was measured by use of two pressure pick-ups set in the line of motion, and the differential pressure was recorded.

3. THE FREE OSCILLATING LIQUID. Now let us consider the free oscillating liquid at increasing frequencies. (Movie 1)* We see the equipment again in a motion picture. There is the tank with its pivot point, the water at a level of 20" above the base of the tank, the shaft, the motor, and the gear.

The tank motion is started. At first, the surface swings like a beam. Except for increase in the amplitude of the water, the motion remains unchanged up to a frequency of about 1.3 cps. Then, suddenly, the liquid begins to slosh, and continues through varying visual phases up to and beyond the natural frequency of the tank, which is 1.43 cps. At a frequency of about 1.7 cps, the sloshing suddenly stops, and the water becomes practically quiescent. This completes the first mode of the natural frequency.

When the differential pressure measurements for the first mode are plotted against frequencies, (Slide 4) it is found that the amplitudes increase up to a certain point. The curve then breaks off, follows an almost horizontal course to another point, after which it slopes downward. The natural frequency is located half way between these two characteristic points. At the first point, sloshing begins, and at the second point, sloshing ends; and both of these points are reproducible.

The curves for other tank sizes indicate the same characteristics. (Slide 5) In the next slide, there are curves for 10", 17.5", and 25" tanks, plotted so that the natural frequencies coincide. The dots are for the 10" tank; the crosses, the 17.5" tank; and the circles, the 25" tank. The abscissa does not represent the absolute frequency, but the difference between the natural frequency and the forced frequency. I should like to consider only the part of the curve up to the natural frequency since this seems the most interesting.

---

* Movies are not reproduced here.

From this curve, it can be seen first, that the curves for different tank sizes coincide and second, that the curves break off a different points, depending on the tank size. The larger the tank diameter, the larger the amplitude during sloshing and the smaller the sloshing range.

The curve itself is a hyperbola with the simple equation

$$xy = \text{constant}$$

The x values in this case are the differences between the frequency of the forced oscillation (f_x) and the natural frequency (f_n). The y values are the pressure amplitudes (a_p). From our experiments, we found the maximum amplitudes at the sloshing point are nearly linearly proportional to the tank diameter. The factor of proportionality was found to be 2π, and we obtained the relation

$$2\pi\, a_p \ (f_n - f_s) = \text{constant}$$

At the sloshing point, this becomes

$$2\pi\, a_{p\ max}\,(f_n - f_s) = d(f_n - f_s) = \text{constant}$$

or $$2\pi\, a_{p\ max} = d$$

However, $2\pi\, a_{p\ max}$ is actually the length of the pressure wave, and this leads to the conclusion that sloshing occurs when the length of the pressure wave becomes larger than the tank diameter.

This result may be considered in regard to the Reynolds number. Generally, the Reynolds number is given as

$$R = \frac{w \cdot \ell}{\nu}$$

Where w = velocity
ℓ = length
ν = kinematic viscosity

In our case

$$w = s\omega_x \qquad \text{and } \ell = d$$

Where s = stroke
ω_x = forced frequency
d = tank diameter

The Reynolds number at natural frequency becomes

$$R_n = \frac{s\omega_n d}{\nu} = \frac{2\pi\, f_n\, sd}{\nu}$$

and the Reynolds number at the sloshing frequency,

$$R_s = \frac{s\omega_s d}{\nu} = \frac{2\pi\, f_s\, sd}{\nu}$$

The difference between the Reynolds numbers is

$$R_n - R_s = \frac{2\pi\, s}{\nu} \cdot d \cdot (f_n - f_s),$$

If the difference in the Reynolds numbers remains the same

then $$d(f_n - f_s) = \left[\frac{(R_n - R_s)\nu}{2\pi s}\right] = \text{constant} = c$$

for equal strokes and viscosities.

In the lower graph, the constant is calculated, using the values given in the upper graph. It establishes a straight line parallel to the abscissa which breaks off and converges to zero at the sloshing frequency.

From these results, it will be seen that sloshing cannot be considered as resonance which takes place only in the immediate vicinity of the natural frequency, producing very high amplitudes. This conception is refuted by the increasing sloshing range for smaller tank diameters and the sudden breaking off of the curve at the point of sloshing. Actually sloshing is a damping effect. Unfortunately for us, the larger the tank diameter, the smaller is this damping effect so that, with large tank diameters, the amplitudes become quite great. From the theoretical formula derived from the equivalent pendulum, the amplitudes are expected to be infinite at natural frequency. The theoretical formula describes typical resonance curves only. These curves are not equal but similar to the pressure curves according to the previous formulae satisfying the experimental results with sufficient approximation. But in all these considerations, there is no explanation for the reduction in amplitude caused by the damping during sloshing.

This brings up the question, "What is the physical reason for the breaking off of the amplitudes during sloshing?" All our tests have shown that there is a certain correlation between the tank size, the frequency and the sloshing point. Furthermore, the observations led to the belief that the sloshing effect is a surface effect almost similar to the surf of the ocean waves which are so-called "surface waves" or "gravity waves." (Slide 6) According to this version the liquid particles on the surface are considered as rotating, with a phase shift depending on the distance from the center of the impact. The surface then forms a wave, with the shape depending on the radius r and the frequency of the rotation $\omega = 2\pi f$. The velocity of the surface wave may be called c and the velocity of the rotating particle $w = \omega r = 2\pi f r$. Then, the velocity of the liquid particles $u_1 = c + w$ at the crest and $u_2 = c - w$ in the trough. Hence, the kinetic energy at the crest is

$$E_1 = \frac{m}{2} u_1^2 = \frac{m}{2} (c + w)^2$$

and in the trough

$$E_2 = \frac{m}{2} u_2^2 = \frac{m}{2} (c - w)^2$$

The difference of these kinetic energies must be equal to the potential energy of the liquid particles at a height difference $h = 2r$,

$$E_1 - E_2 = \frac{m}{2} \cdot \left[(c + w)^2 - (c - w)^2 \right] = \frac{m}{2} \; (4cw) \; = 2m \; g \; r$$

Since $w = 2\pi fr$

$$c = \frac{gr}{w} = \frac{g}{\omega} = \frac{g}{2\pi \; f}$$

and because $c = \lambda \; f$

$$\lambda = \frac{g}{2\pi \; f^2}$$

Let us consider the special case where $c = w$. It is obvious that this is the limit of the stability of the surface wave. The formula

$$c = \frac{gr}{w} \qquad \text{or } cw = gr$$

becomes

$$w^2 = gr$$

or because $w = 2\pi \; fr = \omega r$,

$$\omega^2 r = g$$

and $4\,\pi^2 f^2 r = g$.

Since $2\,\pi f^2 \lambda = g$,

$$\lambda = 2\,\pi\, r$$

This means that the highest amplitude is reached when the acceleration of the rotating particle $\omega^2 r$ is equal to the acceleration due to gravity g. If the particle acceleration becomes greater, the surface wave becomes unstable, and sloshing occurs.

Now, suppose such a surface wave does not proceed to infinity but is stopped by the tank wall and reflected to the opposite wall. Furthermore, suppose the previous development that $\lambda = 2\,\pi\, r$ is satisfied at the sloshing frequency. Then substituting our experimental values for f_s,

$$\lambda = 2d$$

and $$r_{max} = \frac{d}{\pi}$$

The values for the maximum amplitudes calculated in this way concur with the observed values. It is of interest that the maximum amplitude at natural frequency, according to this last formula, becomes

$$r_n = \frac{d}{2\varepsilon}$$

because $\omega^2 r = g$ and $\omega_n^2 = \frac{2\varepsilon g}{d}$

Thus, the amplitude at natural frequency is a little smaller than the amplitude at the point where sloshing begins. This is an excellent agreement with our experience, since in all cases, the recorded pressure is lower at natural frequency than at sloshing point.

Surface wave conditions are treated in detail by Lamb in Chapter IX of "Hydrodynamics". According to Lamb, the velocity w of the rotating particles decreases with the depth according to an e- function, as do the amplitudes for different areas at equal pressures which is demonstrated in the next slide.(Slide 7) The upper curve is a cycloid, and lower curves are trochoids. According to this consideration, sloshing occurs under the conditions where the acceleration of the rotating particles becomes larger than the acceleration due to gravity.

Summarizing, we can state that the behavior of an oscillating liquid can be explained by two different occurrences. The first one forms the normal resonance curve caused by the natural frequency and the second one the sloshing curve. The later gives the limit condition for the stability of the liquid motion within an oscillating container. (Slide 8)

In the next slide all the important relations are plotted. The vertical coordinate indicates the values of the diameters and amplitudes in inches and the horizontal coordinate indicates the frequencies. The curve f_n shows the natural frequencies depending on the diameter. The curves f_{s1}, f_{s5}, f_{s10} show the sloshing frequencies at a stroke of 1", 5", 10", respectively. The curves a_y and a_p indicate the maximum visual amplitudes and the maximum pressure amplitudes. Above these both curves there is sloshing. This is the area of instability and below these curves is the area of stability, where the resonance curves of a 10", 17.5", 25" and 100" tank characterize the condition of the liquid motion depending on the frequency. The crosses indicate measured values. They are in very good agreement with the theoretical curves. Consider, for example, the 25" tank. The natural frequency is 1.2 cps and the sloshing frequency at 1" stroke is 1.1 cps. The maximum visual or level amplitude in the sloshing point is 8" and the corresponding maximum pressure is 4". The curves of the other tank sizes are corresponding. Consequently, with this nomogram the behavior of the liquid in all tank sizes is predictable. This picture shows, too, that immediate conclusions from model tank sizes to original tank sizes are possible with sufficient accuracy.

4. <u>THE HEIGHT OF THE FREE OSCILLATING MASS.</u> In the report by W. Graham and A. M. Rodriguez entitled "The Characteristics of Fuel Motion Which Affect Airplane Dynamics", formulae were developed concerning the relations of the so-called free oscillation mass to the total mass and of the so-called rigid mass to the total mass. From these formulae, it is possible to derive the height of the free oscillating mass as

$$h = 0.26\, d$$

This is valid for a rectangular tank, but it may be expected that the

value is not changed appreciably for a cylindrical tank.

To prove the equation $h = 0.26\ d$ experimentally, we supposed the free oscillating liquid in a U tube (Slide 9) Generally the natural frequency of such a liquid column is

$$\omega_n = \sqrt{\frac{k}{m}}$$

Where k = spring constant = force per unit deflection

And m = mass of free oscillating liquid

If the level in one arm of the tube is raised one inch, the difference in level, h, is 2 inches and the force for restoring the original water level is

$$K = m_h \cdot g = V_h \cdot \rho \cdot g = A \cdot h \cdot \rho \cdot g = 2A \rho g$$

V = volume
ρ = density
A = surface area
g = acceleration due to gravity
ℓ = length

and the total oscillating mass

$$m = V \cdot \rho = A \cdot \ell \cdot \rho$$

Hence $\omega_n^2 = \frac{k}{m} = \frac{2A\rho g}{A\rho\ell} = \frac{2g}{\ell}$

If we were to divide our tank into two parts by a wall in the center, this would be similar to a U tube. At a length , which would represent a certain depth of the wall, the U tube effect may be expected. The mass involved may be considered as a free oscillating mass.

In our case, $\omega_n^2 = \frac{2g\varepsilon}{d} = \frac{2g}{\ell}$

Hence $\ell = \frac{d}{\varepsilon}$

and the depth of the wall $\frac{\ell}{2} = \frac{d}{2\varepsilon} = 0.27d \approx \frac{d}{4}$

This result is in very good agreement with my earlier statement that the depth of the free oscillating mass is about one-fourth of the tank diameter. From the equation for the U tube shown earlier, it can be seen that the surface area and density do not influence the frequency. This lack of dependence on surface area supports our earlier premise that the depth of the free oscillating mass in a cylindrical tank would be the same as that developed by Graham and Rodriguez for a rectangular tank.

The U tube effect can best be demonstrated by the next movie (Movie 2). The wall is in the center of the tank, perpendicular to the motion of the forced oscillation. The tank is oscillated at a constant frequency, and at first, we observe the free oscillating surface. Now, the wall is

2" deep, and the surface swings like a beam. When the wall is lowered to a depth of about one-fourth of the tank diameter, that is 4.5", we get the real U tube effect. When we lower the wall still further, both surfaces oscillate differently, so that the liquid behaves as though we had two separate tanks.

Another interesting observation is the motion of the liquid particles at different depths. (Movie 3) The pattern of the floating body corresponds with our previous conclusion, as the motion of the thread shows decreasing velocity of the liquid particles as the depth is increased, according to an e-function. At a depth of about one-fourth of the tank diameter, the lateral motion of the floating body practically stops.

5. <u>CONCLUSIONS FOR DESIGNING A PROPER DAMPING DEVICE</u>. Our investigations lead to two important conclusions:

First, a free oscillating liquid reaches a maximum amplitude at a frequency lower than the natural frequency, and there is a simple relationship between this maximum amplitude and the tank diameter. With increasing frequency above this point of maximum amplitude, the liquid sloshes, and this sloshing inhibits larger amplitudes. Hence, it is a damping effect.

And second, the depth of the dangerous free oscillating mass is about one-fourth the tank diameter.

From these findings, we can restrict the requirements for damping to a device that will resist the forced motion of the liquid particles in the area equal to a depth of one-fourth the tank diameter.

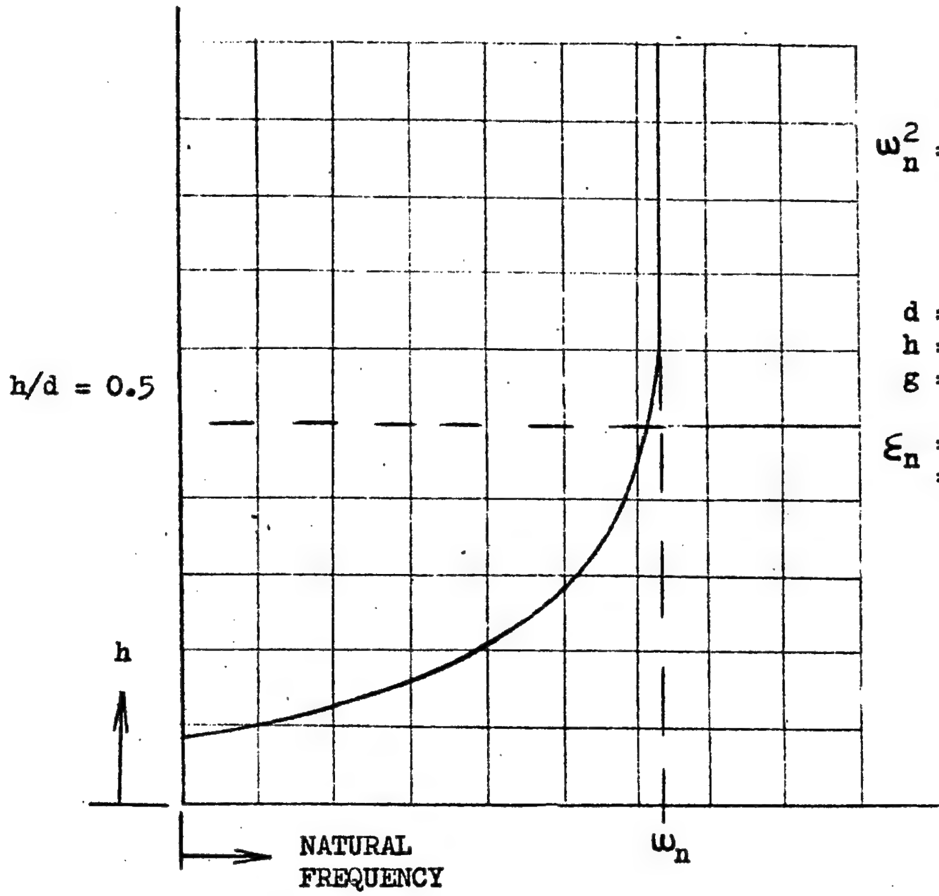
These conditions can be met in a simple manner by a mat consisting of a network of fibers. (Movie 4) The next movie shows such a mat of the same diameter as the tank and of a thickness of about one-fourth the tank diameter. The aluminum balls furnish the bouyancy elements. We oscillate the tank at the sloshing frequency and drop in the mat. The sloshing ceases immediately, while the tank continues oscillating at the sloshing frequency.

In the missile propellant tank, however, the device must adapt itself to the changing surface areas, caused by the pipes and lines running through the tanks. This realization let to the division of the mat into many parts, as shown in the next film. (Movie 5) These cylindrical bodies have a length of one-fourth the tank diameter. The perforated aluminum sleeve is a substitute for the fibre network of the mat, and each body has a floating ball as a buoyancy unit.

The liquid is sloshing strongly, and the float devices are thrown in. The bodies arrange themselves, and when the entire surface area is covered, the liquid comes to rest. Since these devices float, we can empty the tank without any increase in surface amplitude. The liquid remains stabile. Now again, we fill the tank, and the liquid remains calm, despite the fact that throughout this movie the frequency has remained at the sloshing point.

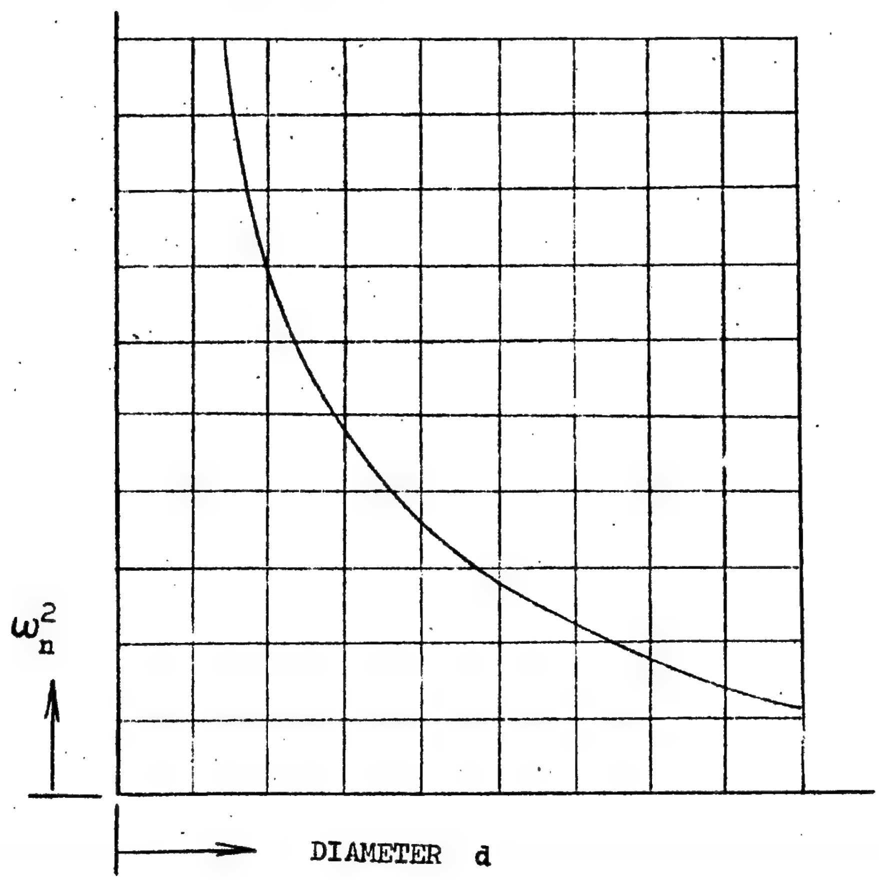
We recognize the device as one possibility for damping sloshing. It follows naturally from the theoretical considerations and certain basic experiments which we have conducted. The next lecture will cover some of the other possibilities. Discussion will be postponed until the conclusion of that paper, when you will have a more complete picture of this investigation.

The available time has been too short to permit more detailed discussion of the principles involved. A report will be issued at ABMA, Huntsville, within a few weeks that will treat these investigations more comprehensively.

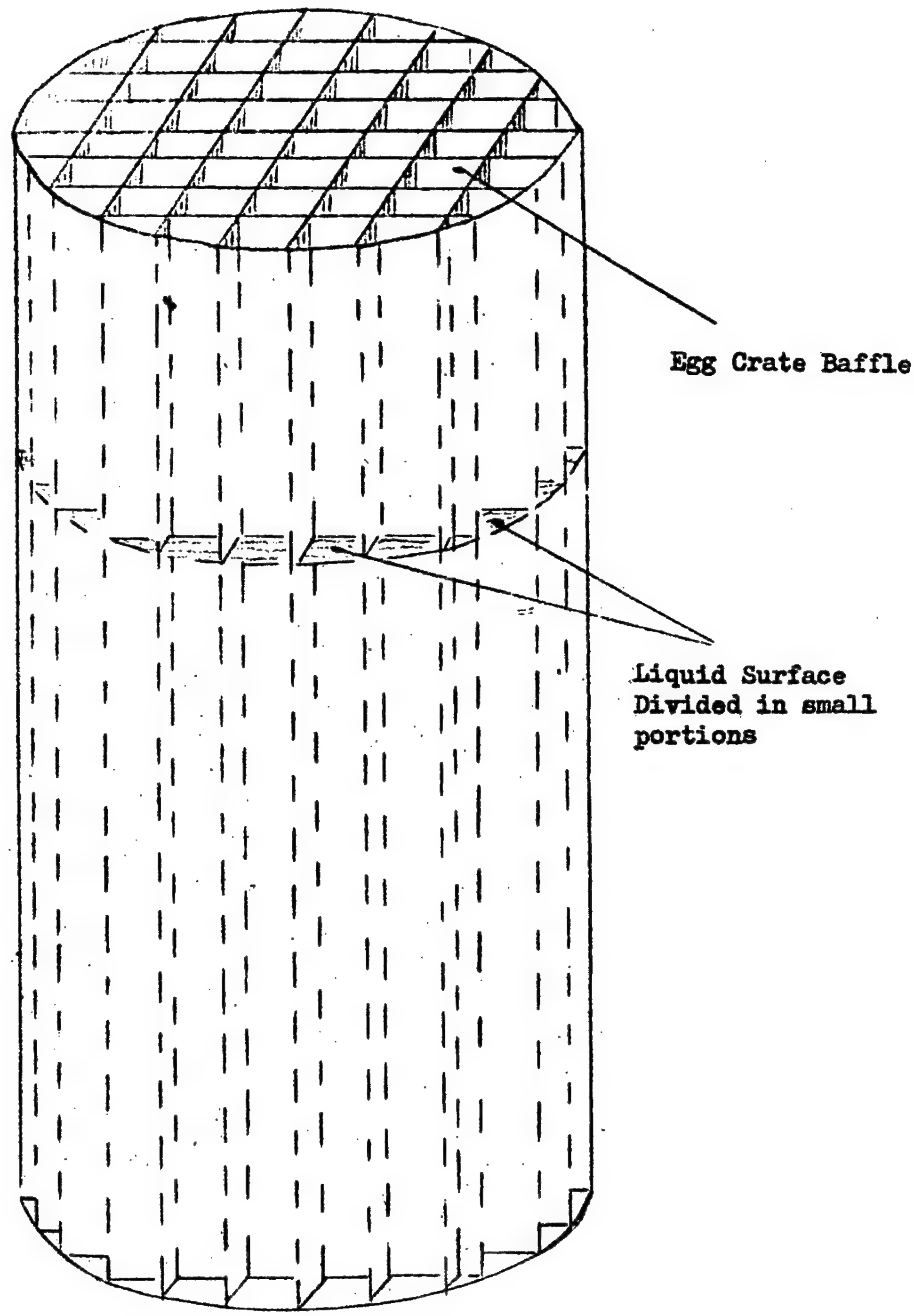


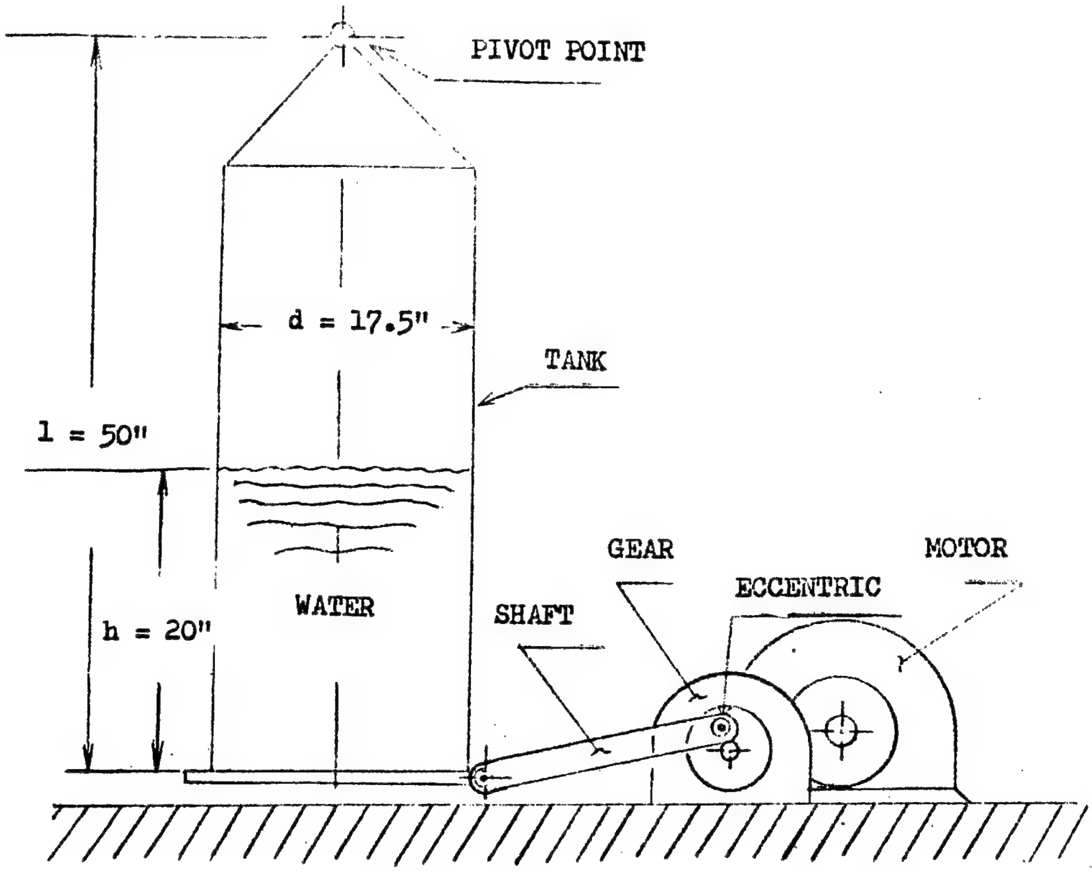
$$\omega_n^2 = \frac{2g\varepsilon_n}{d} \tanh \left(\frac{2\varepsilon_n h}{d}\right)$$

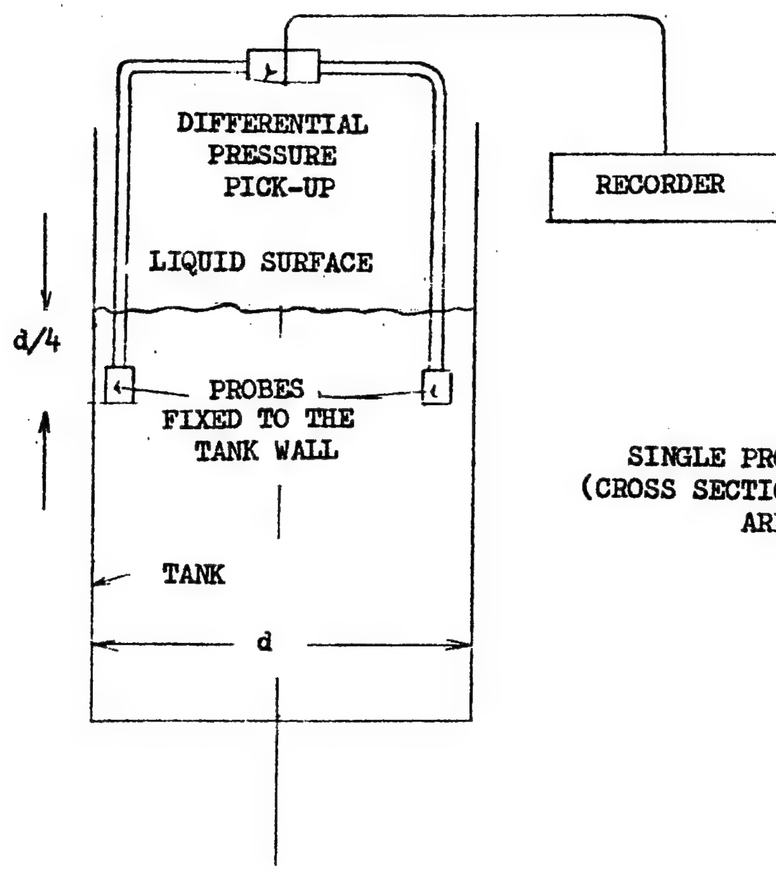
d = TANK DIAMETER
h = HEIGHT OF WATER LEVEL
g = ACCELERATION DUE TO GRAVITY
ε_n = 1.84 (n=1)
= ZERO ε_n OF FIRST DERIVATIVE OF THE BESSEL FUNCTION OF THE FIRST ORDER AND FIRST KIND $F_1^1(\varepsilon_n) =$

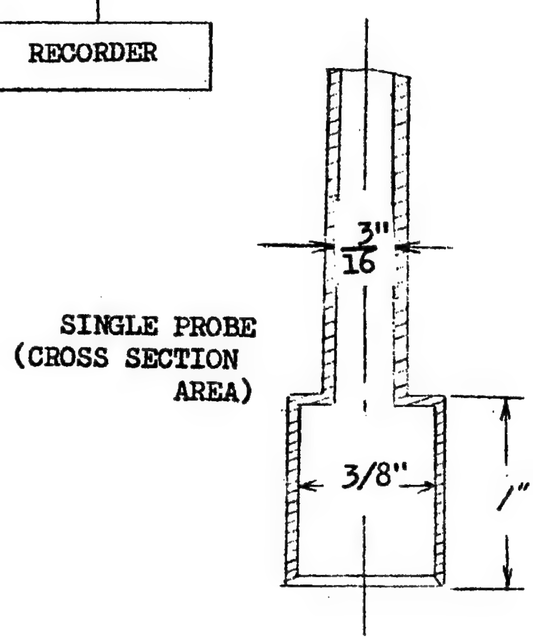


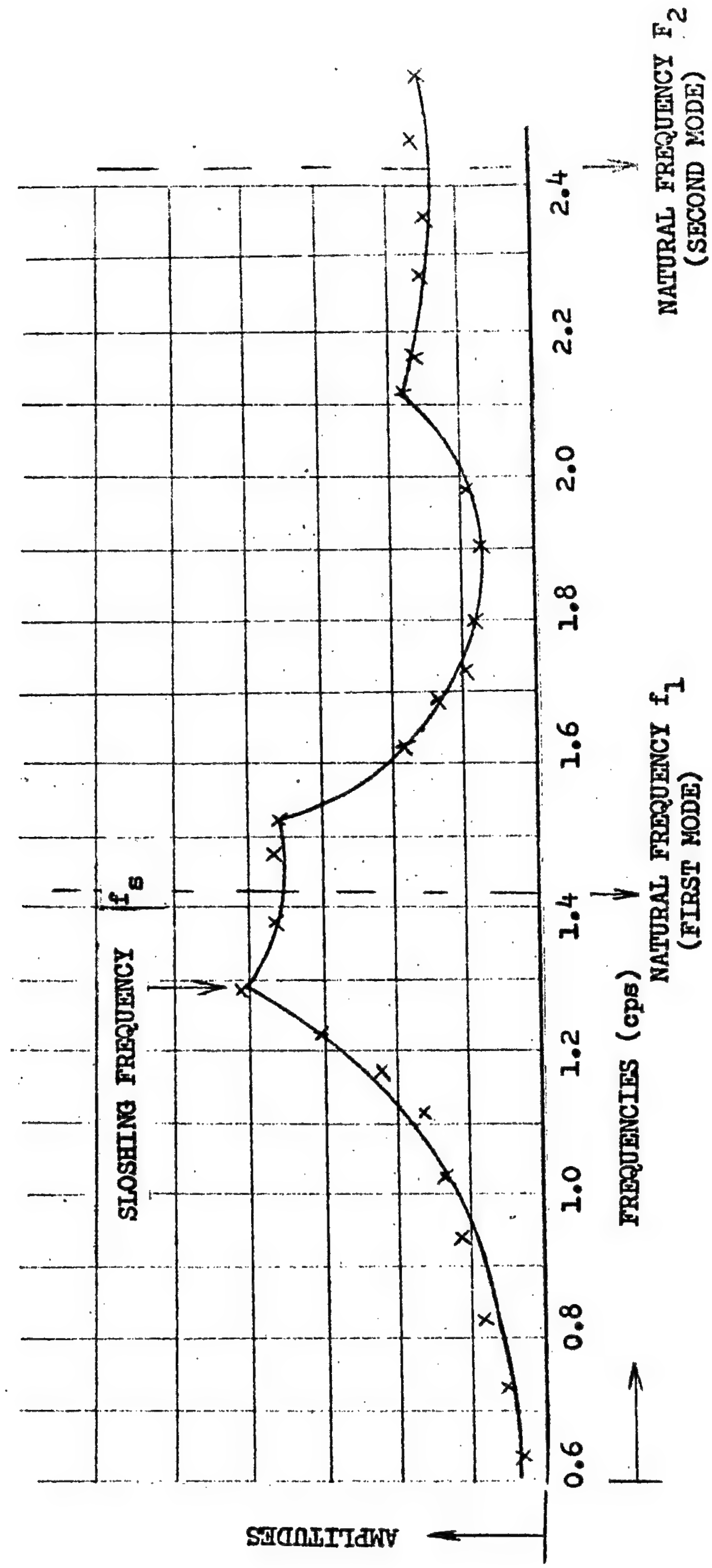
$$\omega_n^2 = \frac{2g\varepsilon_n}{d}$$


Egg Crate Baffle
Liquid Surface
Divided in small
portions

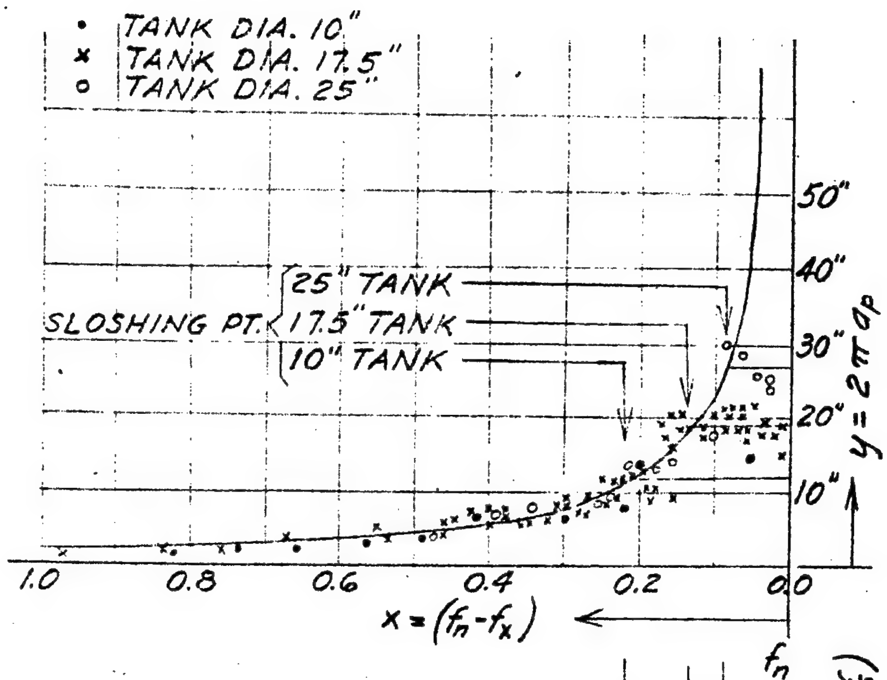

PIVOT POINT
d = 17.5"
TANK
1 = 50"
h = 20"
WATER
GEAR
MOTOR
ECCENTRIC
SHAFT

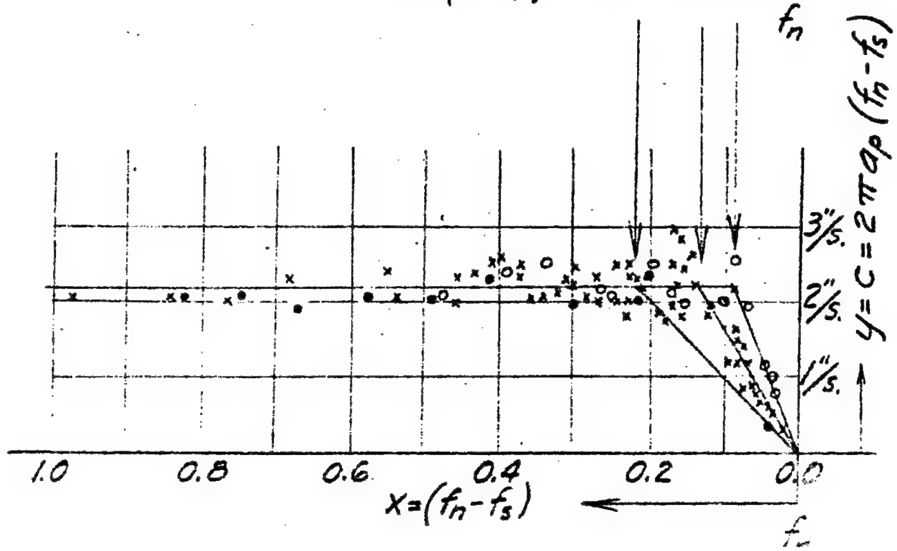

DIFFERENTIAL PRESSURE PICK-UP
RECORDER
LIQUID SURFACE
d/4
PROBES FIXED TO THE TANK WALL
TANK
d


3"/16
SINGLE PROBE (CROSS SECTION AREA)
3/8"
1"



SLOSHING FREQUENCY
f_s
NATURAL FREQUENCY f_1
(FIRST MODE)
NATURAL FREQUENCY F_2
(SECOND MODE)
FREQUENCIES (cps)
AMPLITUDES
0.6
0.8
1.0
1.2
1.4
1.6
1.8
2.0
2.2
2.4





$$xy = 2\pi ap\ (f_n - f_x) = \text{CONSTANT}$$

$$2\pi ap\max\ (f_n - f_s) = d(f_n - f_s)\ \text{CONSTANT}$$

$$2\pi ap\max = d$$

$$R = \frac{W-L}{V} \qquad W = S - \omega_x \qquad L = d$$

$$R_n = \frac{S\omega_n d}{V} = \frac{2\pi s f_n d}{V}$$

$$R_s = \frac{S\omega_s d}{V} = \frac{2\pi s f_s d}{V}$$

$$R_n - R_s = \frac{2\pi s}{V} \cdot d(f_n - f_s)$$

$$d(f_n - f_s) = \frac{R_n - R)V}{2\pi s} = \text{CONSTANT} = C = 2.3 \text{ in/sec}$$

R = REYNOLDS NUMBER

W = VELOCITY

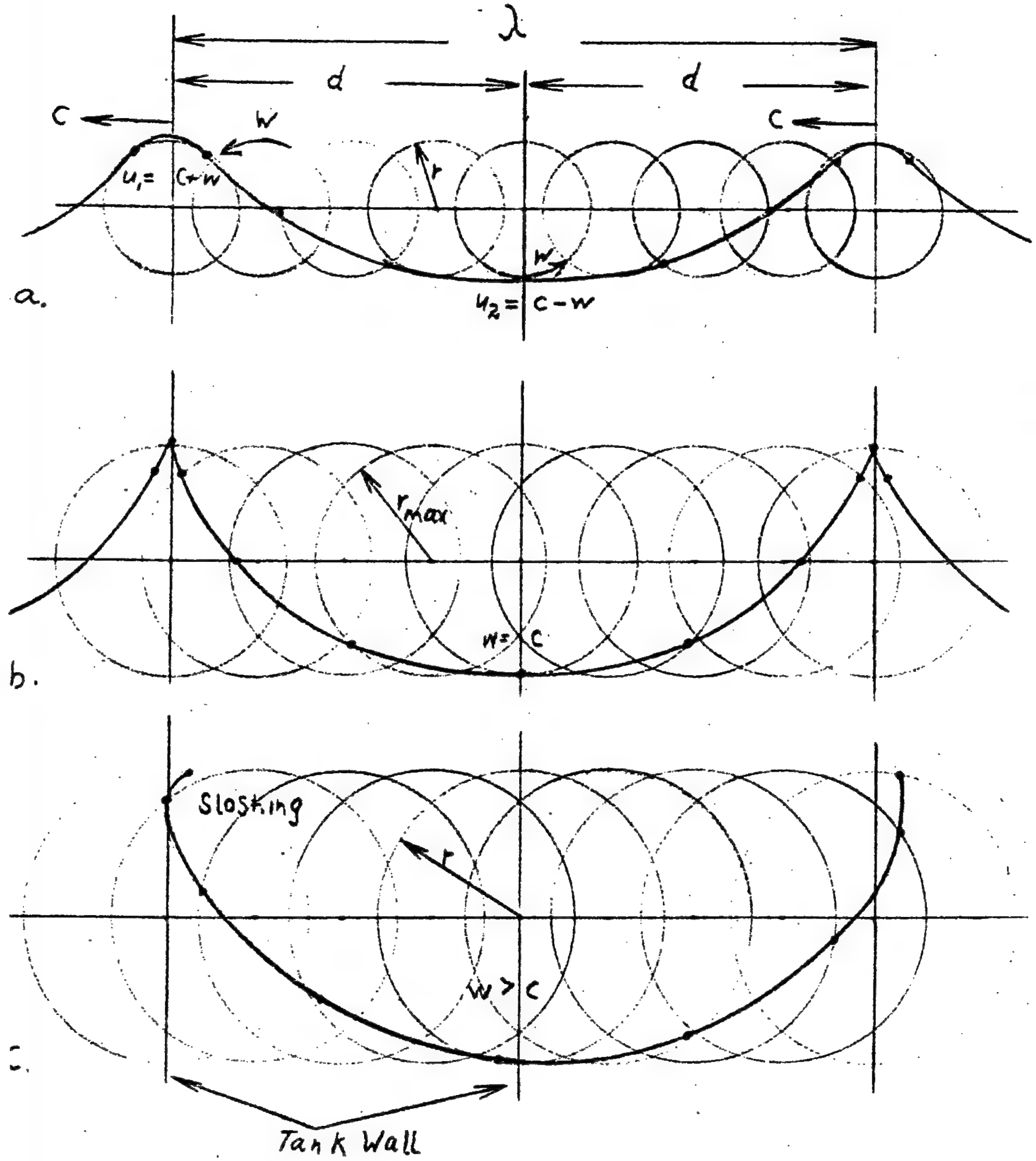
L = LENGTH

S = STROKE

= 2 f = FREQUENCY

d = TANK DIAMETER

V = KINEMATIC VELOCITY



c = velocity of the surface wave

w = velocity of the rotating particles

u_1= velocity at the crest

u_2= velocity at the trough

$u_1 = c + w \qquad w = 2r\pi f = \omega r$

$u_2 = c - w$

a. $E_1 = \frac{m}{2}(c + w)^2 \quad E_2 = \frac{m}{2}(c - w)^2$

$E_1 - E_2 = 2mcw = \underbrace{2mgr}_{\text{potential energy between crest and trough}}$

$c = \frac{gr}{w} = \frac{g}{\omega} = \frac{g}{2\pi f} = \lambda f$

$\lambda = \frac{g}{2\pi f^2}$

---

b. $w = c$

$w^2 = gr$

$\omega^2 r = g$

$4\pi^2 f^2 r = g = 2\pi f^2$

$\lambda = 2\pi$

$\lambda = 2d$

If $\omega_n^2 r = g$

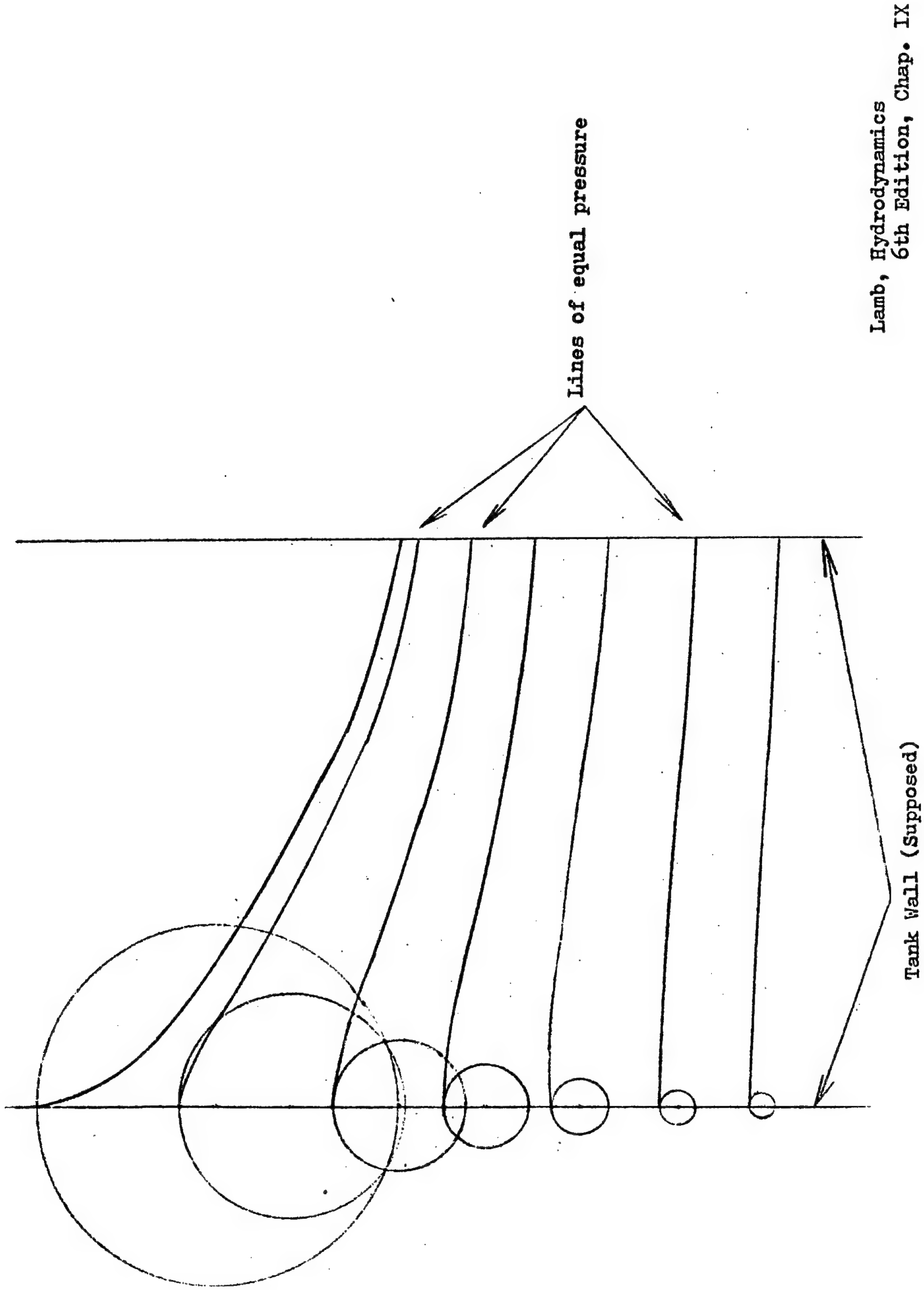
then $g = \frac{2\varepsilon g r}{d}$

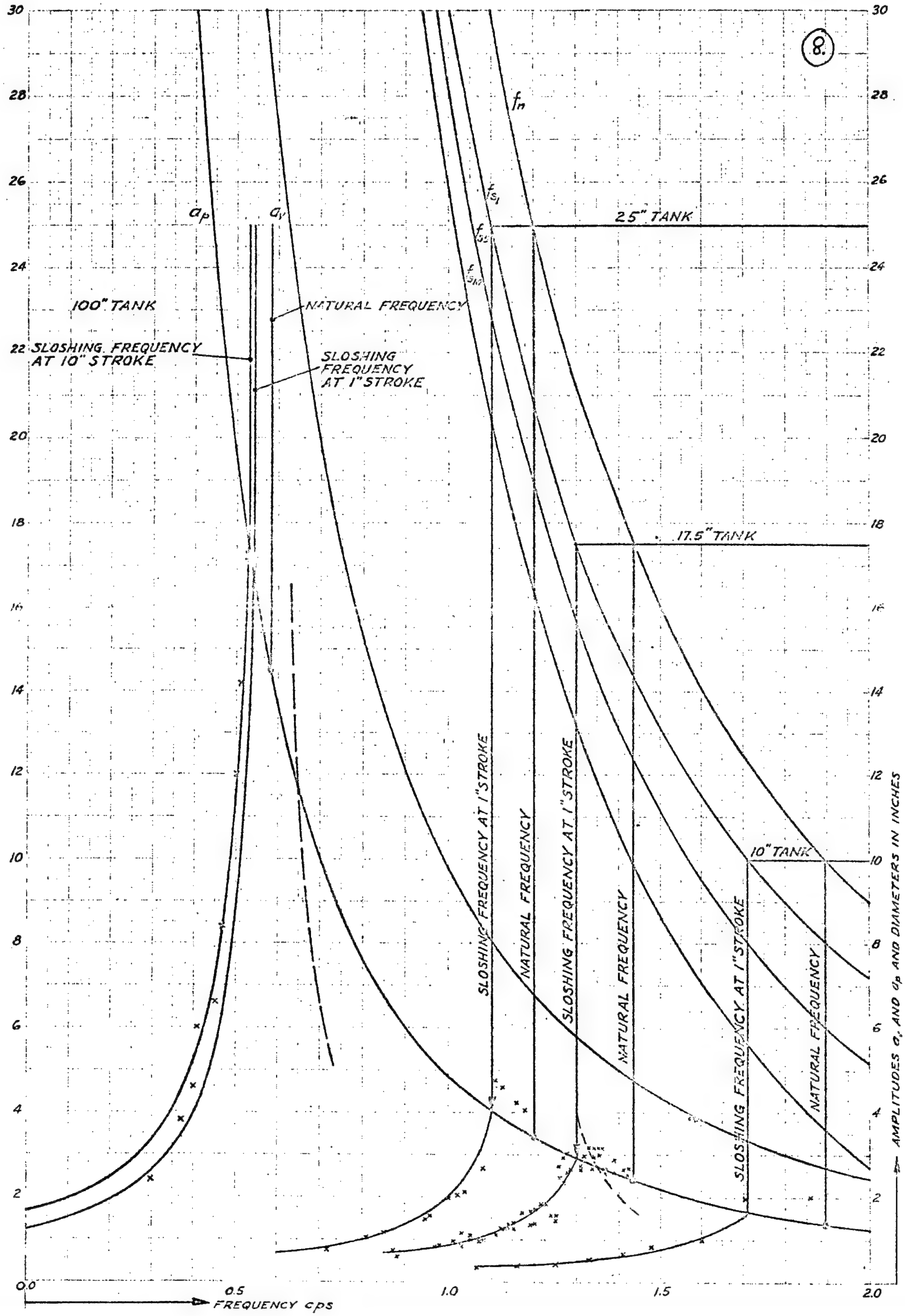
because

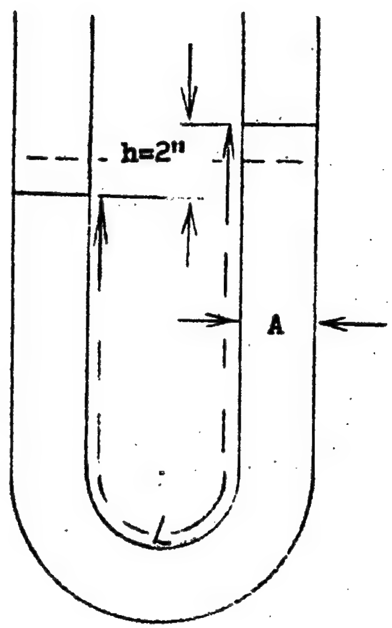
$\omega_n^2 = \frac{2\varepsilon g}{d}$ for $\tanh \frac{2\varepsilon h}{d} = 1$

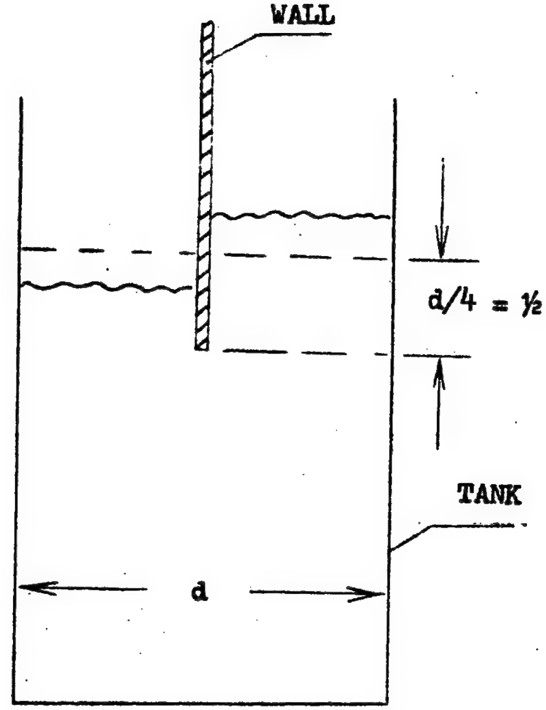
therefore

$r_s = \frac{d}{\pi} \qquad r_n = \frac{d}{2\varepsilon} \qquad \varepsilon = 1.84$


Lines of equal pressure
Tank Wall (Supposed)
Lamb, Hydrodynamics
6th Edition, Chap. IX


8.
100" TANK
SLOSHING FREQUENCY AT 10" STROKE
NATURAL FREQUENCY
SLOSHING FREQUENCY AT 1" STROKE
a_p
a_v
f_n
f_{s_1}
f_{s_5}
$f_{s_{10}}$
25" TANK
17.5" TANK
10" TANK
SLOSHING FREQUENCY AT 1" STROKE
NATURAL FREQUENCY
SLOSHING FREQUENCY AT 1" STROKE
NATURAL FREQUENCY
SLOSHING FREQUENCY AT 1" STROKE
NATURAL FREQUENCY
AMPLITUDES a_v AND a_p AND DIAMETERS IN INCHES
FREQUENCY cps
0.0
0.5
1.0
1.5
2.0
2
4
6
8
10
12
14
16
18
20
22
24
26
28
30





$$\omega_n = \sqrt{\frac{K}{m}}$$

$$K = m_n \cdot g = V_n \rho g = A \cdot h \cdot \rho g = 2A \cdot \rho \cdot g$$

$$m = V \cdot \rho = A \cdot L \cdot \rho$$

$$\omega_n^2 = \frac{K}{m} = \frac{2A \cdot \rho \cdot g}{A \cdot \rho \cdot L} = \frac{2g}{L}$$

$$\omega_n^2 = \frac{2g\varepsilon}{d} = \frac{2g}{L}$$

$$L = \frac{d}{\varepsilon}$$

$$\frac{L}{2} = \frac{d}{2\varepsilon} \approx \frac{d}{4}$$

K = FORCE PER UNIT DEFLECTION
m = MASS = $V \cdot \rho$
V = VOLUME
A = SURFACE AREA
ρ = MASS PER CUBIC INCH
g = ACCELERATION DUE TO GRAVITY

Part II

DESIGN OF ANTI SLOSH DEVICES AND PERFORMANCE OF EXPERIMENTS

A Ballistic Missile is supposed to follow the path of a Keplerian ellipse in its free flight phase. It is the task of the Guidance and Control System to bring the actual trajectory in coincidence with the reference trajectory, that means into the Keplerian ellipse, at a minimum of errors in order to assure a target hit. (Slide 1)*

There are known-forces and unknown-forces acting on the missile during its powered flight. It is most desirable to continuously constrain the missile as closely as possible to the reference trajectory, in order to keep deviations and necessary corrections as small as possible.

Large liquid propellant Ballistic Missiles have tanks of immense dimensions; they are subjected to oscillations and accelerations in their powered flight phase. The designer is facing the problem of suppressing the violent motions of the propellants in the missile tanks in order to keep the unknown-forces at a minimum.

In the tanks, when the wave amplitudes, agitated by oscillations, approach breaking height, the state of "sloshing is reached. (Slide 2) (Still picture of Sloshing).

This sloshing of the liquids exerts considerable forces on the missile structure, affects the controls and renders any liquid level measuring device inaccurate, if not impossible.

The larger the missile, the greater the danger that missile oscillations agitate sloshing in the propellant tanks, that means a sloshing damping device has to be provided. (Slide 3)

What are the requirements for such a device?

1. A device that will produce a high damping effect.

2. A device that will absorb the slosh forces or transfer the forces to the tank structure uniformly, not creating points of stress concentrations.

3. A device that will not change the moment of inertia of the liquid when under roll-oscillations.

4. A device that will divide the free oscillating liquid mass into partial masses so that all separated portions cannot oscillate freely and their surface diameters are relatively small.

5. A device encompassing and covering the entire cross-sectional area.

6. A device filling the free oscillating zone of the liquid or filling a height ¼ the diameter of the tank beneath the liquid surface.

---

* Slide can be found at the end of Part II.

7. A device that will follow the lowering of the liquid level without interrupting the damping effect.

8. A device adaptable to changes in the shape of the surface of the cross-sectional area of the tank configuration, not clinging or sticking to pipes or other elements in the tank structure.

9. A device of minimum weight.

10. A device that will utilize little space.

11. A device that will produce no adverse effects on the emptying of the container.

12. A device that will function at high or low temperatures or at other variables in the state of the fluid or environs.

13. A device that is easy to assemble.

14. A device that will not interfere with entry to the inside of the tank for cleaning purposes.

15. A device that will not cause damage to the tank or other build-in equipment during transportation of the missile.

Quite a number of proposals for sloshing suppression have been discussed, designed and tested. (Slides 4 to 11)

1. Devices fixed to the tank structure. (Slide for each type)

   a. Concentric tubular baffles. (1 cylinder, and 2 cyclinders)

      (1) Solid

      (2) Perforated

   b. Cross Baffles (minimum of egg-crating case)

      (1) Solid

      (2) Perforated

   c. Conical ring baffles, 45° upright.

      (1) Solid

      (2) Perforated

   d. Conical ring baffles, 45°, inverted

      (1) Solid

      (2) Perforated

e. Solid conical ring baffles, inverted - Perforated conical ring baffles, upright.

f. Accordion type baffles

(1) Solid

(2) Perforated

2. Devices, floating on the liquid surface

a. Bell - Type Float

b. Mat - Type Float (Slide for each type)

c. Can - Type Float

The first requirement requested was a good damping effect of the Anti Slosh Device. Here arises the question what method can be used to judge the damping effect.

We started out in taking slow motion pictures of the liquid surface and tried to measure and evaluate the visual amplitudes of the liquid surface. By computing the ratio of Wave Amplitudes of the depressed liquid to Wave Amplitudes of the free liquid a damping efficiency factor might be established.

Since the unstable state of motion of the surface (Slide 12) does not permit an accurate measurement of wave amplitudes this method was omitted or used only as a quick means to decide if a proposed device is bearing merits for further detailed considerations.

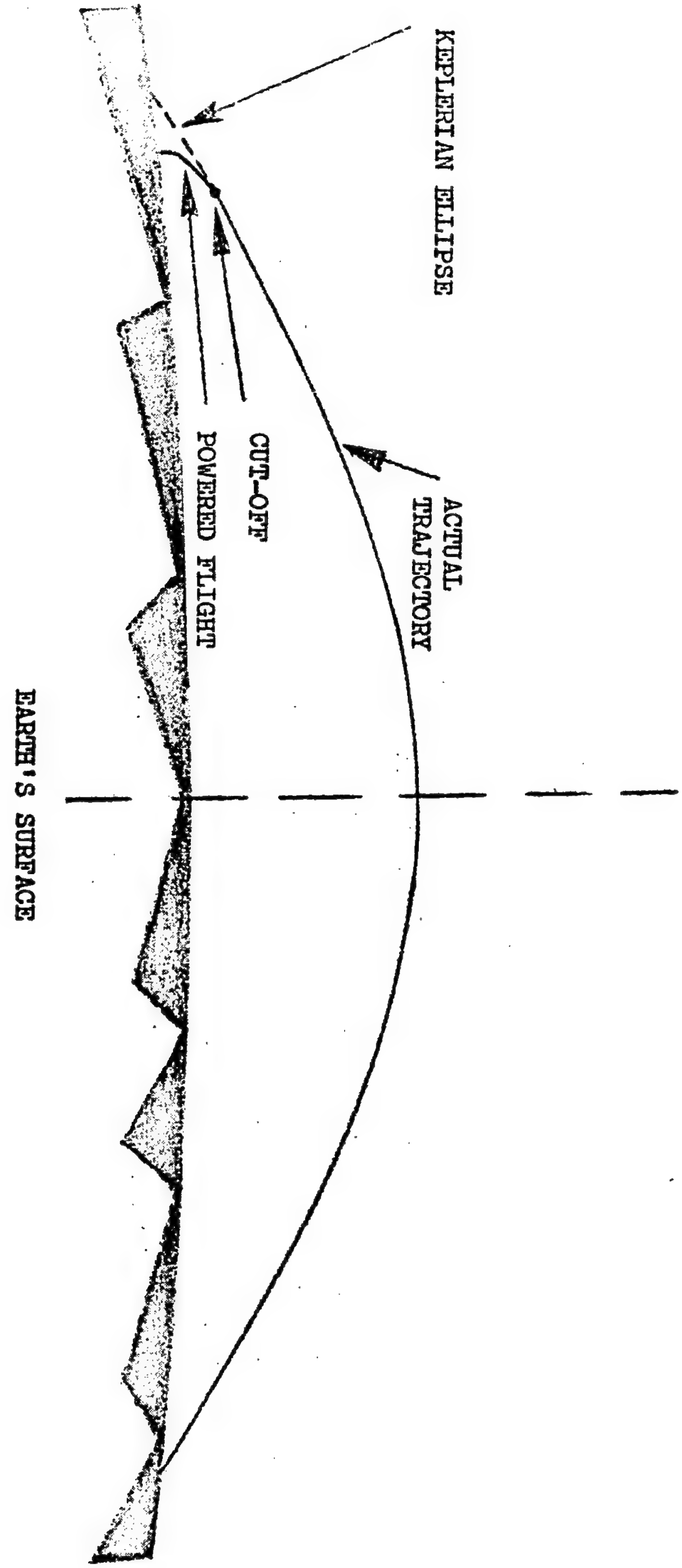
Pressure probe measurements have the disadvantage of not covering a larger surface area; therefore, pressure probe measurements were also ruled out for efficiency considerations.

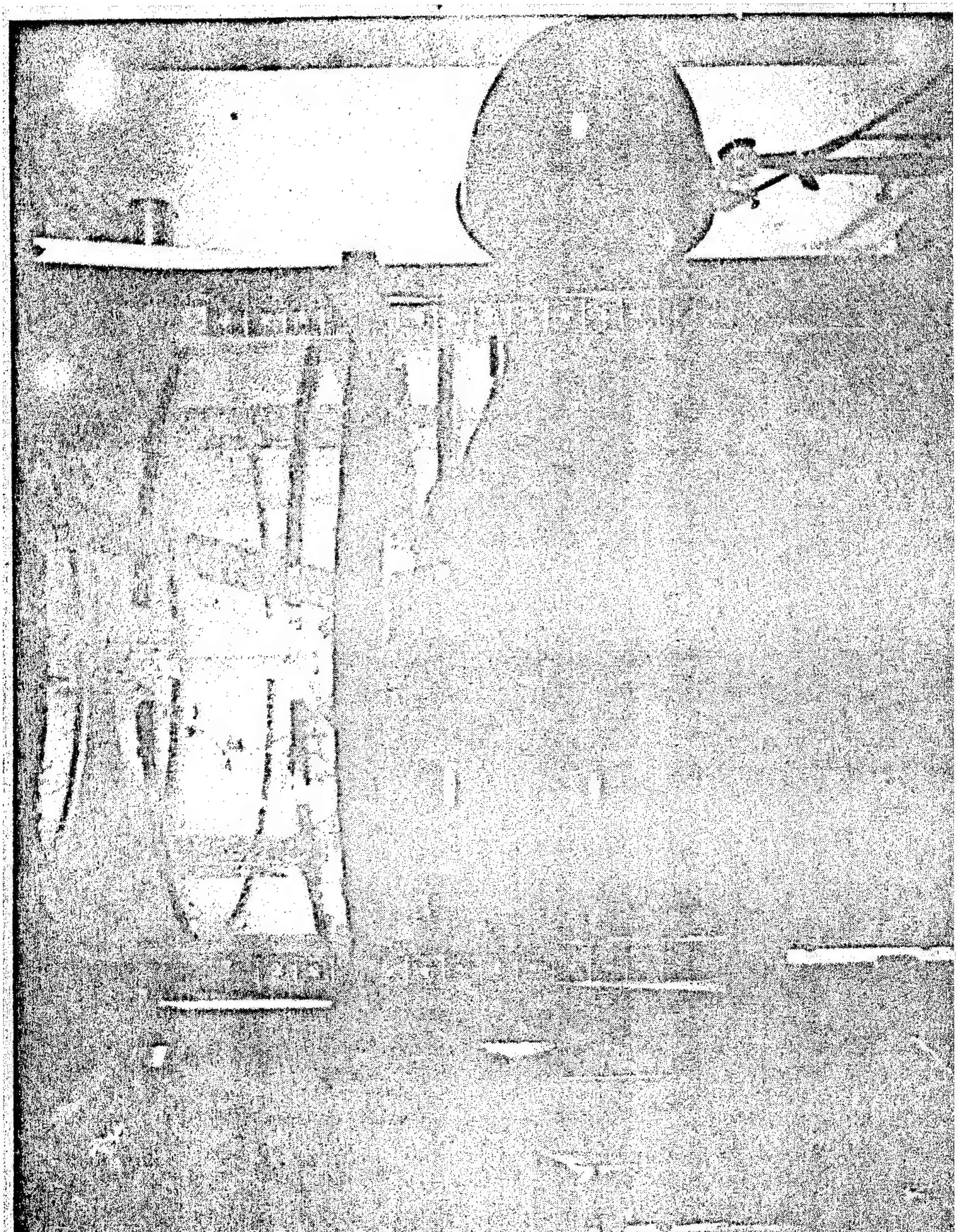
The force necessary to move the tank and thereby agitating the liquid in the tank was finally used as the best indicating value to judge the damping effect of a device. (Slide 13)

Plotting the force as function of mass and acceleration we receive a parabula. The force-curve of the completely restricted liquid, for example, enclosed by a tight lid on the surface is identical with this parabula and would be the curve of ideal damping. (Slide 14)

The next slide (15) is showing the force-curves of devices, which gave the best results in damping the sloshing.

Film Strip on Slosh Damping Devices.

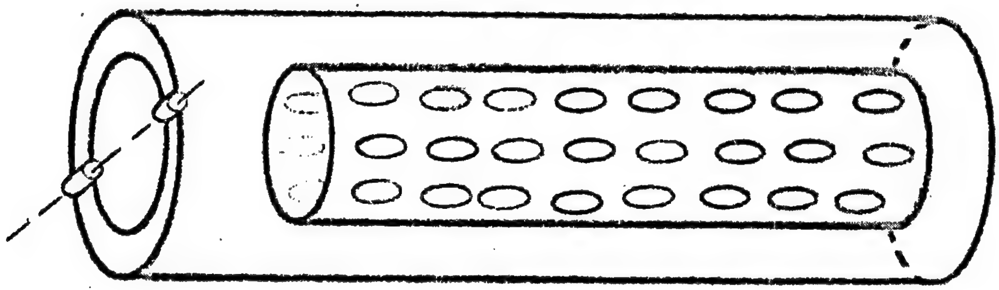

KEPLERIAN ELLIPSE
CUT-OFF
POWERED FLIGHT
ACTUAL TRAJECTORY
EARTH'S SURFACE



ANTI-SLOSH DEVICE

WHAT ARE THE REQUIREMENTS FOR SUCH A DEVICE?

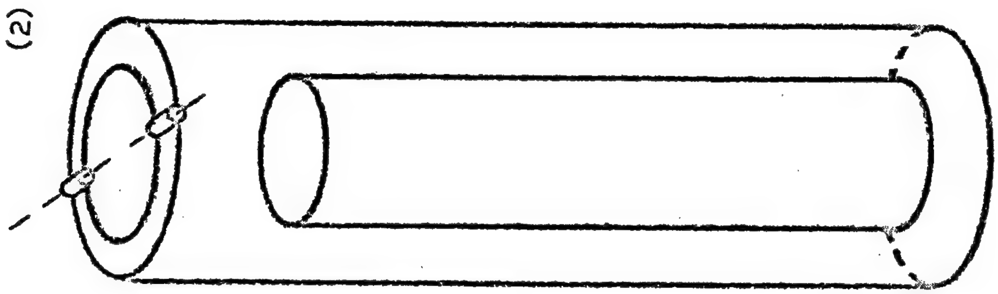
1. A DEVICE THAT WILL PRODUCE A HIGH DAMPING EFFECT.
2. A DEVICE THAT WILL ABSORB THE SLOSH FORCES OR TRANSFER THE FORCES TO THE TANK STRUCTURE UNIFORMLY, NOT CREATING POINTS OF STRESS CONCENTRATIONS.
3. A DEVICE THAT WILL NOT CHANGE THE MOMENT OF INERTIA OF THE LIQUID WHEN UNDER ROLL-OSCILLATIONS.
4. A DEVICE THAT WILL DIVIDE THE FREE OSCILLATING LIQUID MASS INTO PARTIAL MASS SO THAT ALL SEPARATED PORTIONS CANNOT OSCILLATE FREELY AND THEIR SURFACE DIAMETERS ARE RELATIVELY SMALL.
5. A DEVICE ENCOMPASSING AND COVERING THE ENTIRE CROSS-SECTION AREA.
6. A DEVICE FILLING THE FREE OSCILLATING ZONE OF THE LUQUID OR FILLING A HEIGHT ¼ THE DIAMETER OF THE TANK BENEATH THE LIQUID SURFACE.
7. A DEVICE THAT WILL FOLLOW THE LOWERING OF THE LIQUID LEVEL WITHOUT INTERRUPTING THE DAMPING EFFECT.
8. A DEVICE ADAPTABLE TO CHANGES IN THE SHAPE OF THE SURFACE OF THE CROSS-SECTION AREA OF THE TANK CONFIGURATION, NOT CLINGING OR STICKING TO PIPES OR OTHER ELEMENTS IN THE TANK STRUCTURE.
9. A DEVICE OF MINIMUM WEIGHT.
10. A DEVICE THAT WILL UTILIZE LITTLE SPACE.
11. A DEVICE THAT WILL PRODUCE NO ADVERSE EFFECTS ON THE EMPTYING OF THE CONTAINER.
12. A DEVICE THAT WILL FUNCTION AT HIGH OR LOW TEMPERATURES OR AT OTHER VARIABLES IN THE STATE OF THE FLUID OR ENVIRONS.
13. A DEVICE THAT IS EASY TO ASSEMBLE.
14. A DEVICE THAT WILL NOT INTERFERE WITH ENTRY TO THE INSIDE OF THE TANK FOR CLEANING PURPOSES.
15. A DEVICE THAT WILL NOT CAUSE DAMAGE TO THE TANK OR OTHER BUILT-IN EQUIPMENT DURING TRANSPORTATION OF THE MISSILE.

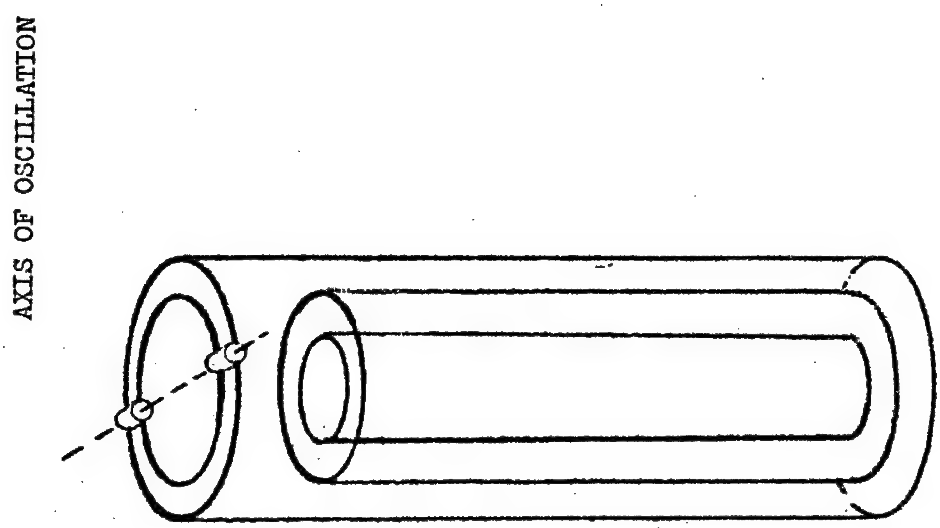


CONCENTRIC TUBULAR BAFFLES (1 CYLINDER)

(1) SOLID

(2) PERFORATED

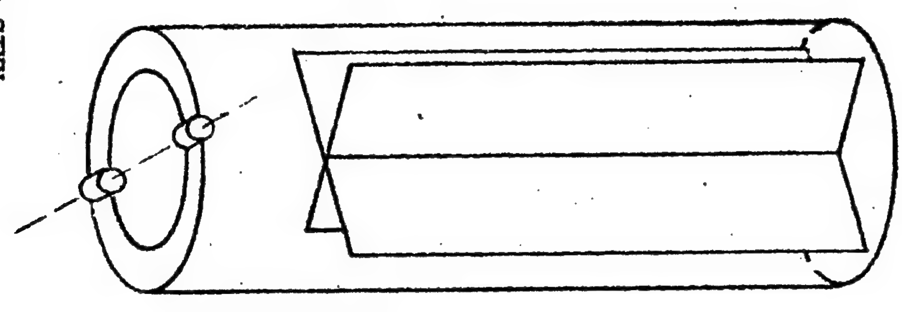




CONCENTRIC TUBULAR BAFFLE (2 CYLINDER)

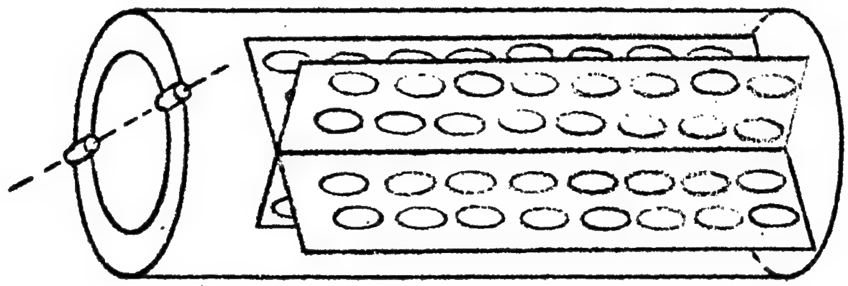
CROSS BAFFLES, MINIMUM OF EGG-CRATING CASE
(1) SOLID
(2) PERFORATED (LARGE HOLES)

AXIS OF OSCILLATION



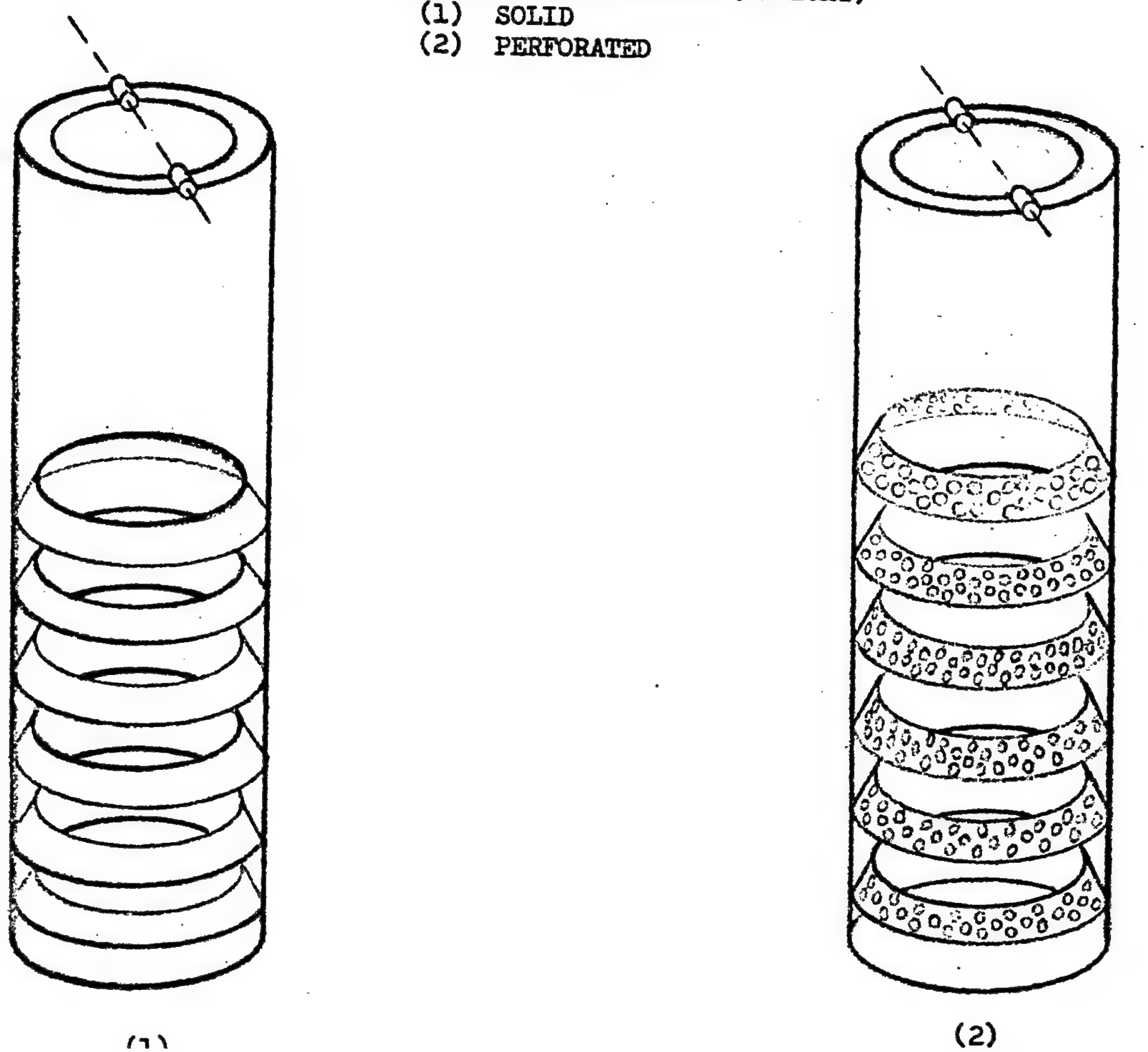
1

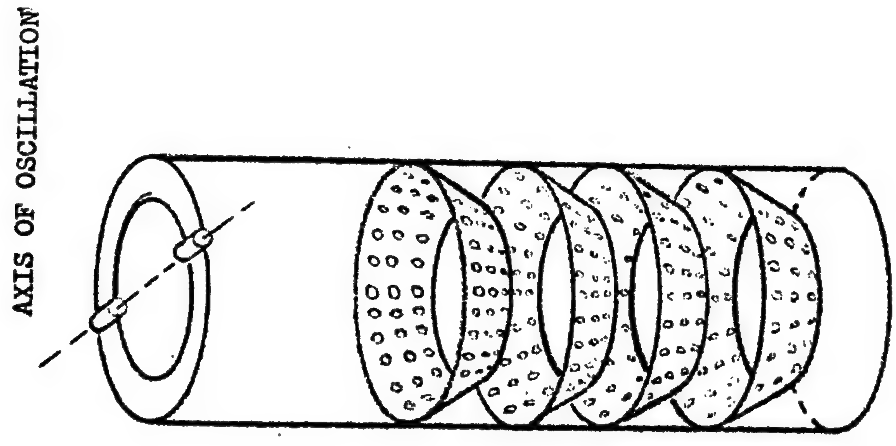
AXIS OF OSCILLATION



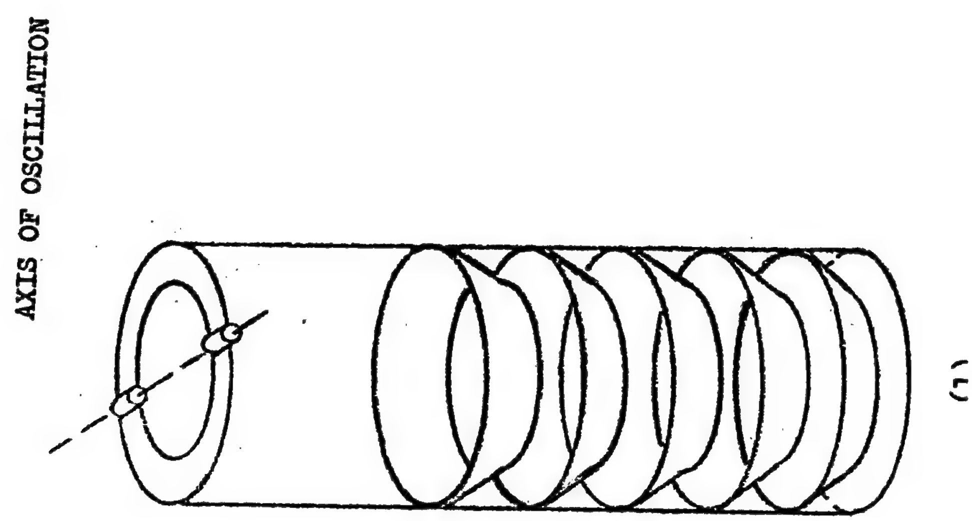
2

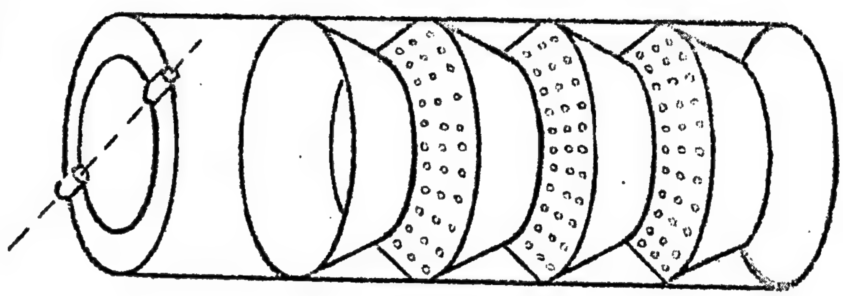
CONICAL RING BAFFLES (UPRIGHT)
(1) SOLID
(2) PERFORATED

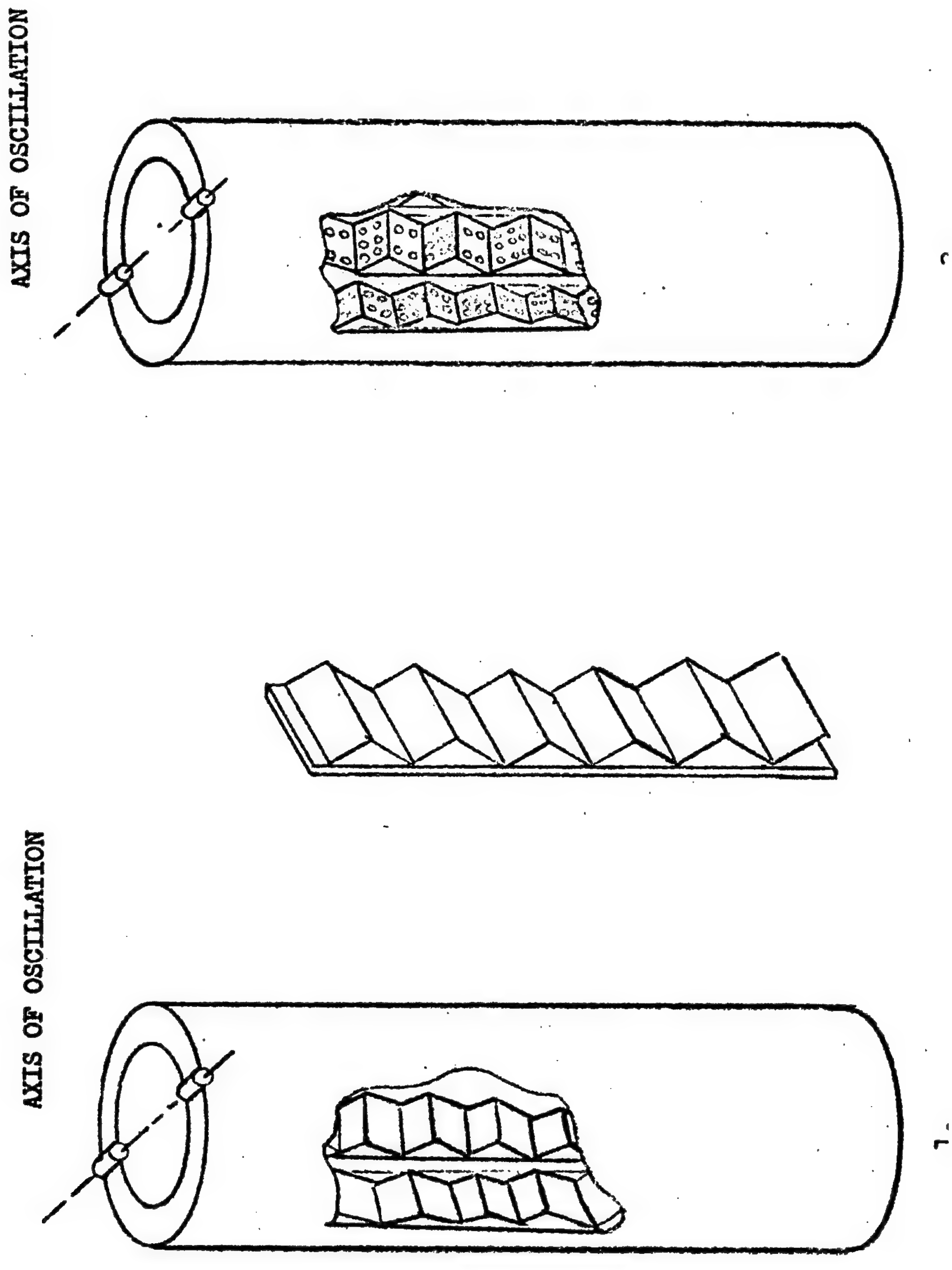




CONICAL RING BAFFLES, INVERTED
(1) SOLID
(2) PERFORATED




AXIS OF OSCILLATION
SOLID CONICAL RING BAFFLES, INVERTED
PERFORATED CONICAL RING BAFFLES, UPRIGHT



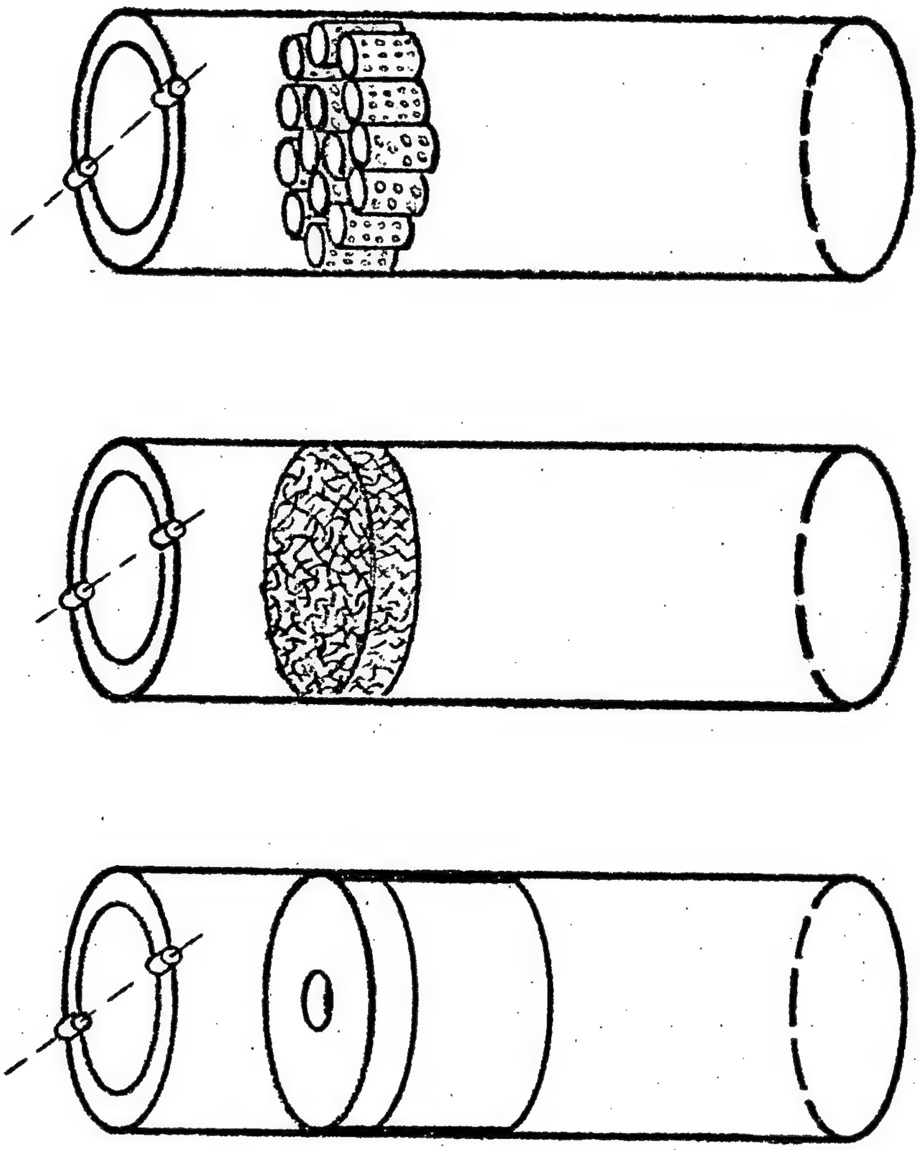
ACCORDION TYPE BAFFLES
1. SOLID
2. PERFORATED

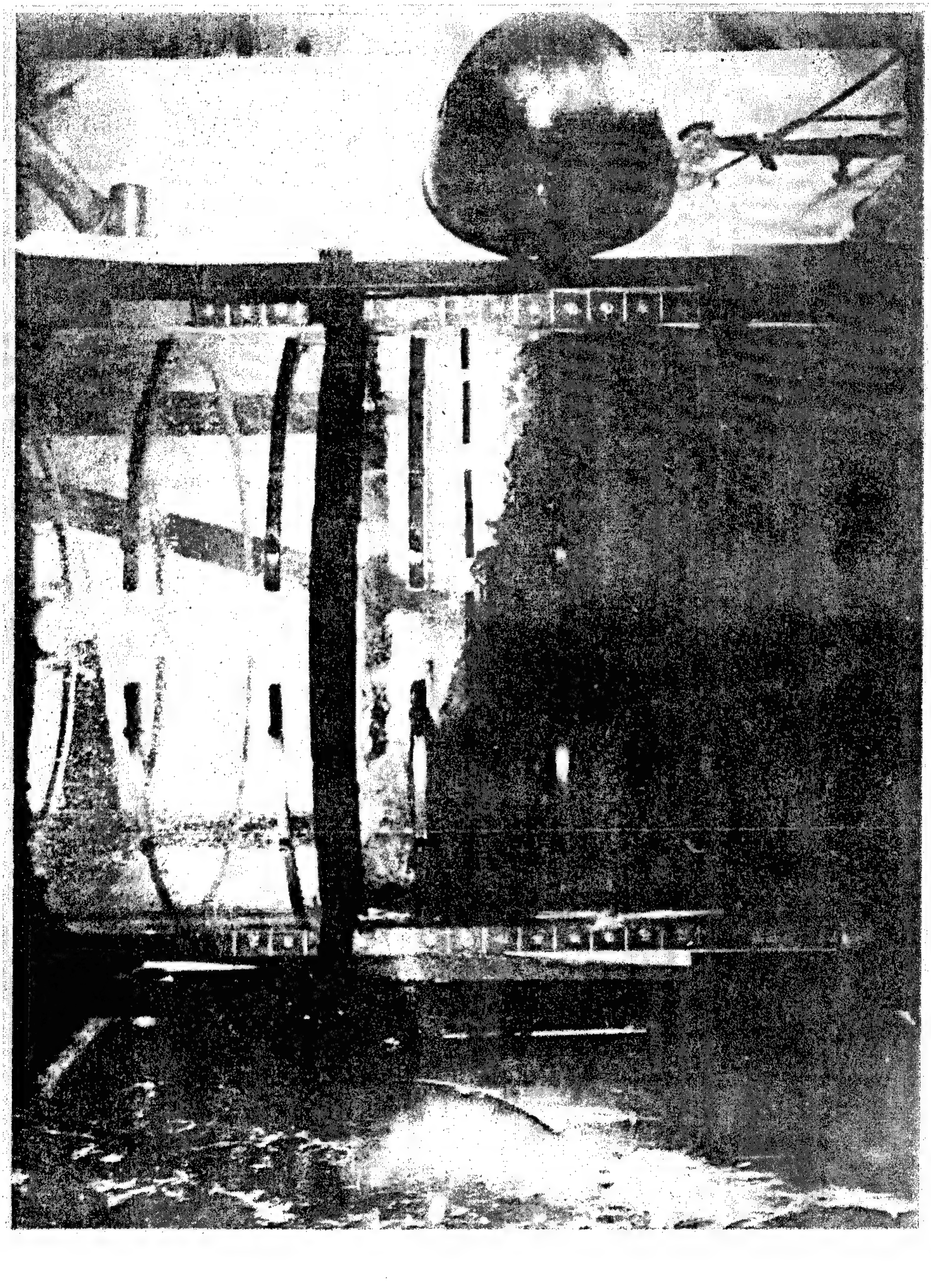
DEVICES FLOATING ON THE LIQUID SURFACE

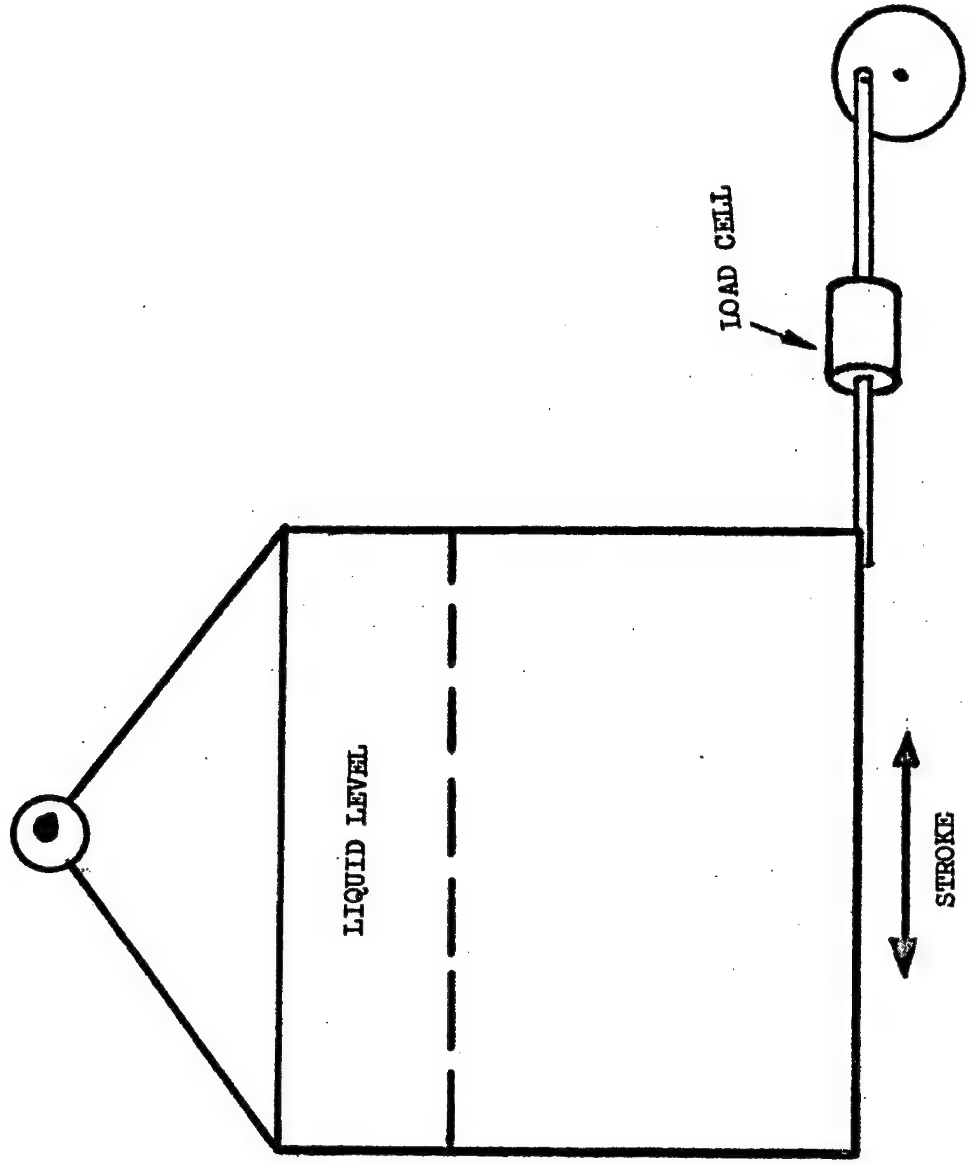
A. BELL TYPE FLOAT

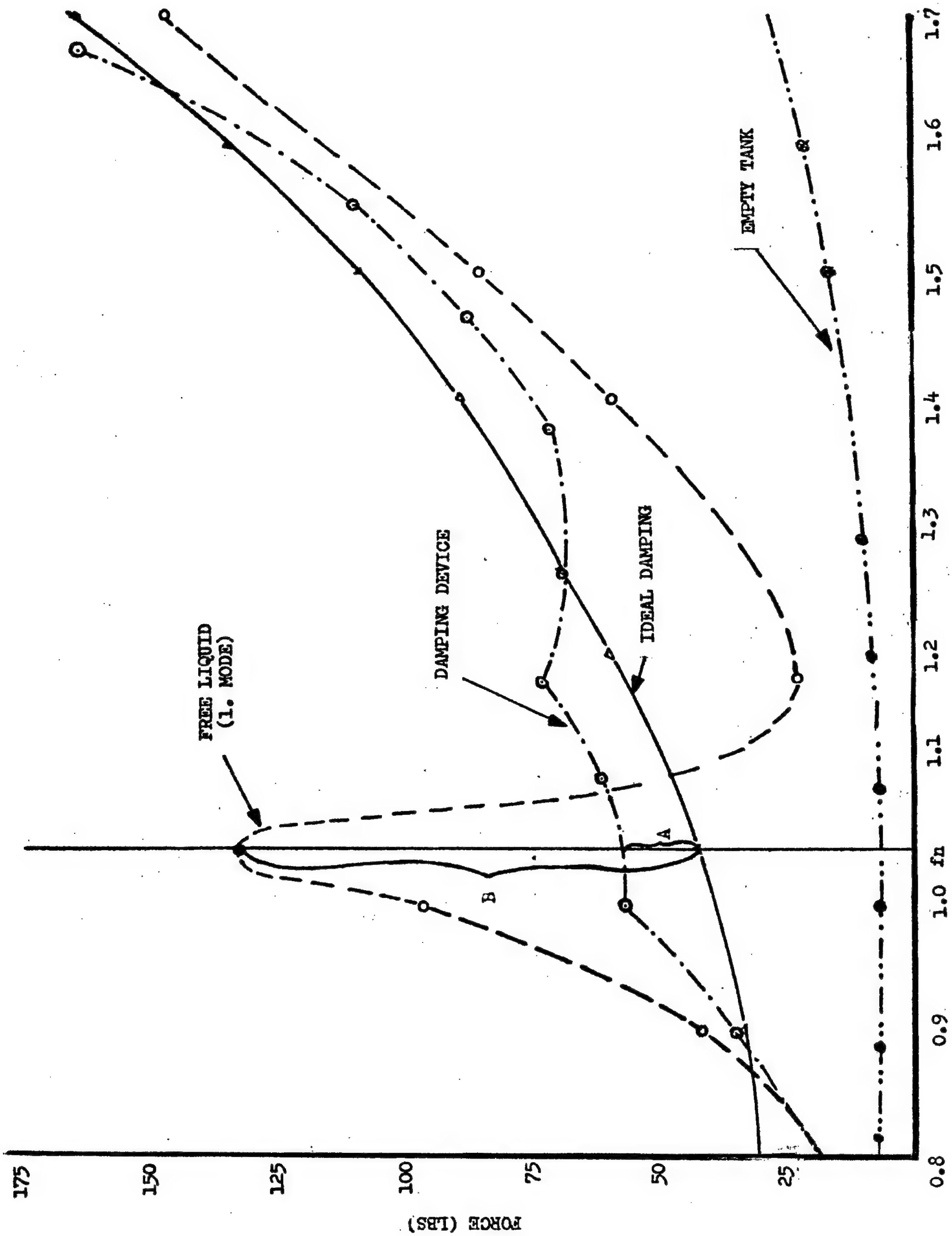
B. MAT TYPE FLOAT

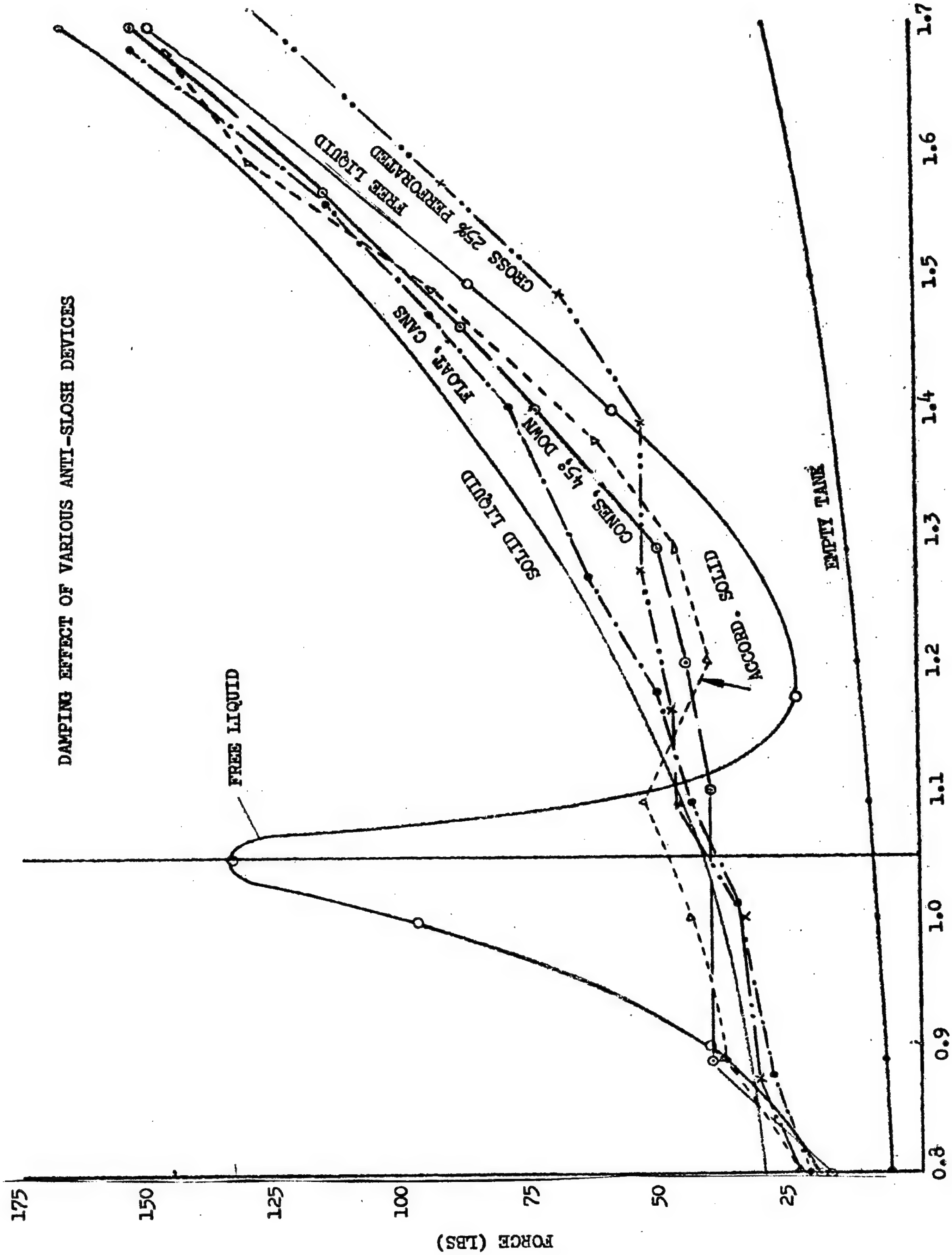
C. CAN TYPE FLOAT






LOAD CELL
LIQUID LEVEL
STROKE


EMPTY TANK
IDEAL DAMPING
DAMPING DEVICE
FREE LIQUID
(1. MODE)
A
B
FORCE (LBS)
175
150
125
100
75
50
25
0.8
0.9
1.0
fn
1.1
1.2
1.3
1.4
1.5
1.6
1.7


DAMPING EFFECT OF VARIOUS ANTI-SLOSH DEVICES
FREE LIQUID
CROSS 25% PERFORATED
FREE LIQUID
FLOAT, CANS
CONES, 45° DOWN
SOLID LIQUID
ACCORD. SOLID
EMPTY TANK
FORCE (LBS)
175
150
125
100
75
50
25
0.8
0.9
1.0
1.1
1.2
1.3
1.4
1.5
1.6
1.7

THE ANALYSIS OF TEST DATA FOR THE PURPOSE OF SETTING SPECIFICATION LIMITS*

P. G. Sanders
Army Rocket and Guided Missile Agency

INTRODUCTION. The purpose of this paper is to discuss some of the problems in choosing acceptance limits for the acceptance testing of cast double-base solid propellant rockets.

Because it is difficult to relate the performance parameters measured in static tests to the actual flight performance of the rocket, (that is, determining unacceptable values of static measurements from their effort on range, time of flight or acceleration), it has often been necessary to use earlier static test acceptance data of production lots that performed satisfactorily in flight to define acceptable values of the static measurements.

In order to fix ideas, I shall describe briefly a typical situation in which the above problem arises. Double-base rocket motors are produced from a base grain powder that is made up in large batches. The totality of motors produced from a single base grain powder batch is called a "Base Grain Lot". Since it is desirable to accept or reject motors in arbitrary lots of smaller size than a Base Grain Lot, a sampling unit of smaller size is given in the specifications and is referred to as a "Motor Lot". Usually, the producer is given a choice of several motor lot sizes, the larger lot sizes having a lower sampling rate. It is left to the producer to balance the cost of the increased sampling rate for smaller motor lots against the undesirability of having a large inventory of untested motors on hand. A typical base grain lot may consist of 10 motor lots, each consisting of approximately 100 motors. From each motor lot a random sample of a specified number of motors is taken and static fired under carefully controlled conditions. The results of these very expensive static tests serve both as a basis for acceptance-rejection by the purchaser and as quality control for the producer. Less expensive small scale tests are conducted, but the full motor tests are considered necessary. The instrumentation for static testing varies among different installations. In all cases the motor is mounted against a load cell and fired in place. The load cell provides a record of thrust versus time; other instruments record motor pressures over time. Some installations provide duplicate channels for all records. From these records various quantities of interest, such as action time, max pressure, average thrust, and total impulse (the integral of thrust over action time) are reduced. The relative importance of these quantities depends on the particular weapons system in which the motor will be used. For simplicity in the following, only one quantity, total impulse, is considered, but the same considerations may apply to other quantities. The data used for illustration are fictitious. They are intended only to illustrate the methods used. The situation is further idealized

*This paper appeared on the program of the Third Conference on the Design of Experiments under the joint authorship of P. G. Sanders and Boyd Harshbarger.

by assuming that an equal number of motor lots is produced from each base grain lot. The different numbers of motor lots per base grain lot present no problems for the estimation of the variance components, but they require more approximations in using the components to set limits. The sampling rate is small enough to make finite population correction factors negligible.

The problem then, is to select limits for total impulse such that if the mean of the values for the sample fired from a motor lot does not fall between them, the motor lot will be rejected. These limits must be such that a negligible amount of production similar to past acceptable production will be rejected, while any detected change in level of total impulse will be cause for rejection of the lot.

<u>SOLUTION</u>. The solution to this problem has been to obtain unbiased estimates of the variance components arising at various stages in sampling. From these, the variance of the mean of the sample from a motor lot is estimated. Then Satterthwaite's approximate degree of freedom method* is used to calculate limits of the form,

$$\mu \pm t\left(\frac{\alpha}{2}, f'\right) \hat{\sigma}_{\bar{x}}$$

where μ is the mean of past acceptable static testing, $\wedge$ denotes estimate and $t\left(\frac{\alpha}{2}, f'\right)$ is a student's "t" value based on f' degrees of freedom and the desired confidence level. f' is the approximate degrees of freedom of $\hat{\sigma}_{\bar{x}}^2$

The sampling scheme is depicted in Figure 1.** A typical analysis of variance for this situation is given in Figure 2, in which are also given estimates of the variance components due to instrumentation (σ_I^2) motor variation within a motor lot (σ_M^2), motor lot variation (σ_{ML}^2) and finally base grain variation (σ_{BG}^2). On the basis of this analysis, and many others, we accept $\sigma_{ML}^2 = 0$.

The true variance of a Motor Lot mean in future sampling will be

$$\sigma_{\bar{x}}^2 = \frac{\sigma_{\cdot}^2}{4} + \frac{\sigma_I^2}{2} + \frac{\sigma_{BG}^2}{1},$$

---

*See Reference on last page.

**Figures can be found at the end of this article.

which may be estimated by

$$\hat{\sigma}^2_{\bar{x}} = \frac{1}{40}\left(9\Delta_2^2 + \Delta_4^2\right)$$

$$= 1.45 \times 10^6$$

In general, $\hat{\sigma}^2_{\bar{x}}$ is not good enough estimate of $\sigma^2_{\bar{x}}$ to justify using it with normal curve area tables to calculate specification limits with a given probability of accepting good material. Neither can we use it with a student's "t" distribution, because the number of degrees of freedom associated with it is unknown, and it does not meet all the assumptions for use with that distribution. However, using an approximation due to Satterthwaite*, we may calculate an approximate degrees of freedom, f', by the following formula:

$$f' = \frac{\left(\hat{\sigma}^2_{\bar{x}}\right)^2}{\left(\frac{9\Delta_2^2}{40}\right)^2 \frac{1}{f_2} + \left(\frac{\Delta_4^2}{40}\right)^2 \frac{1}{f_4}}$$

Observe that the coefficients of Δ_2^2 and Δ_4^2 are the coefficients in the linear combination of then which equals $\hat{\sigma}^2_{\bar{x}}$. For our data

$$f' = \frac{\left(1.45 \times 10^6\right)^2}{\left(\frac{9 \times 5 \times 10^6}{40}\right)^2 \frac{1}{60} + \left(\frac{13 \times 10^6}{40}\right)^2 \frac{1}{5}}$$

$$= 50$$

If we are willing to reject 100 α percent of good motors (α will be very small) then we look up $t\left(\frac{\alpha}{2}, f'\right)$, the 100(1 - α) percentage point of the students "t" distribution for f' degrees of freedom. The acceptance limits are then:

upper limit: $\mu + t\left(\frac{\alpha}{2}, f'\right)\hat{\sigma}_{\bar{x}}$

lower limit: $\mu - t\left(\frac{\alpha}{2}, f'\right)\hat{\sigma}_{\bar{x}}$

For our sample problem we find for $\alpha = .001$,

upper limit: $\mu + t\left(\frac{.001}{2}, 50\right)\hat{\sigma}_{\bar{x}}$

$\mu + 3.50 \times 1.20 \times 10^3$

$\mu + 4200$

---

*See Reference on last page

lower limit:

$$\mu - 4200$$

This completes the solution of the problem.

ADDITIONAL INFORMATION AVAILABLE. The analysis of variance (Figure 2) contains valuable information about the relative magnitudes of instrumentation errors, σ_I^2, and true motor to motor variability, σ_M^2. Techniques are available* which allow exact confidnece intervals to be placed on the ratio $\frac{\sigma_M^2}{\sigma_I^2}$. However, it should be remembered that $\hat{\sigma}_I^2$ measures only differences between channels on a given firing and does not measure how these channels may drift with time. Thus, $\hat{\sigma}_I^2$ may be regarded as an estimate of a lower limit for the variance of all the instrumentation errors, while $\hat{\sigma}_M^2$ may be seriously overestimated. Cognizance of this condition may be necessary when trying to determine whether to spend research effort to reduce σ_I^2 or to reduce σ_M^2.

REFERENCE

"Statistical Theory in Research" by R. L. Anderson and T. A. Bancroft, McGraw Hill Book Company, Inc., 1952.

FIGURE 1: SAMPLING SCHEME FOR ACCUMULATED ACCEPTANCE TESTING DATA

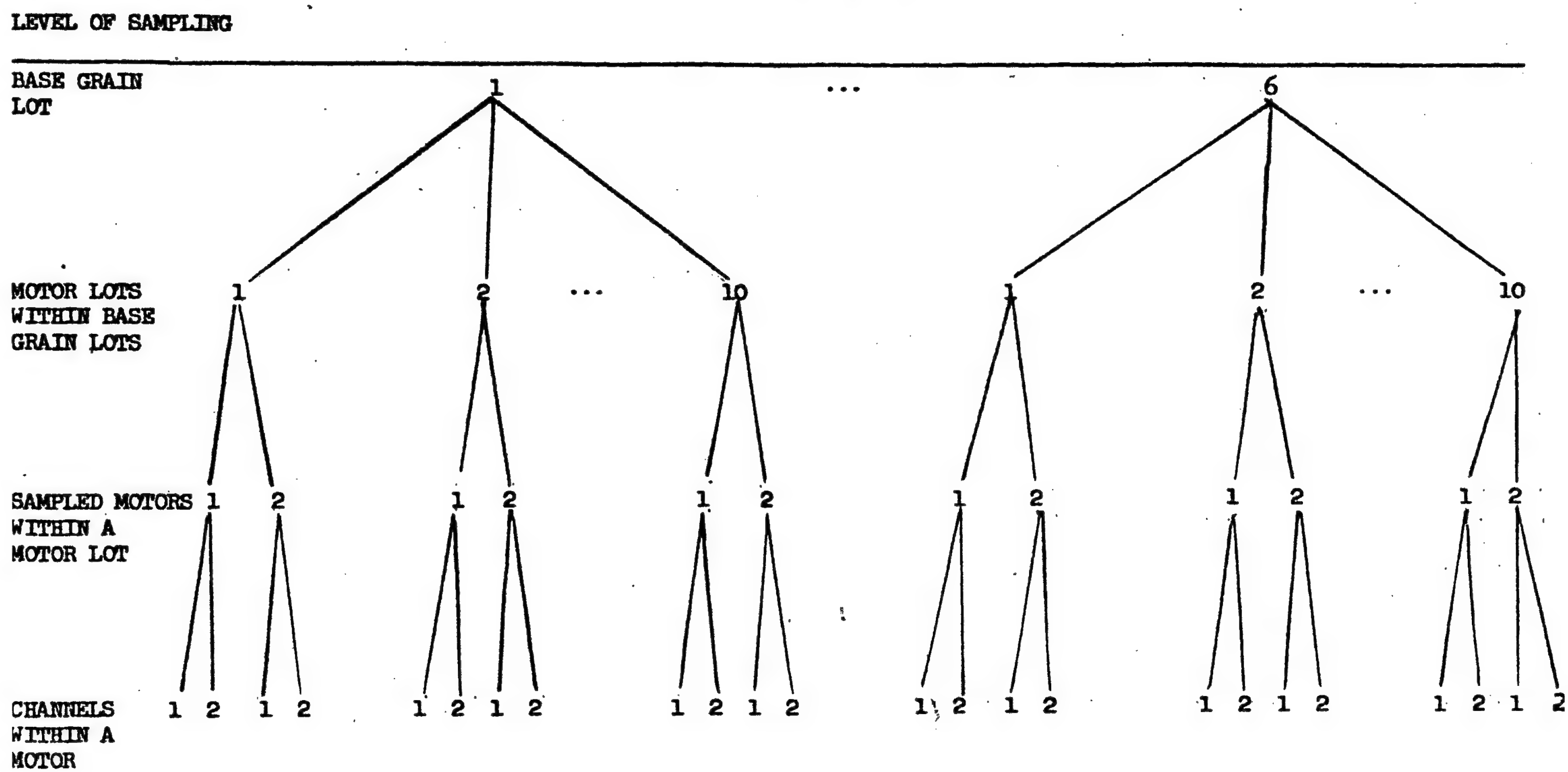


FIGURE 2: ANALYSIS OF VARIANCE AND ESTIMATES OF VARIANCE COMPONENTS

(a) ANALYSIS OF VARIANCE

SOURCE	DEGREES OF FREEDOM	MEAN SQUARE	EXPECTED VALUE OF MEAN SQUARE
Between Base Grain Lots	$f_4 = 5$	$s_4^2 = 13.0 \times 10^6$	$\sigma_I^2 + 2\sigma_M^2 + 4\sigma_{ML}^2 + 4\sigma_{BG}^2$
Between Motor Lots within Base Grain Lots	$f_3 = 54$	$s_3^2 = 4.0 \times 10^6$	$\sigma_I^2 + 2\sigma_M^2 + 4\sigma_{ML}^2$
Between Motors within Motor Lots	$f_2 = 60$	$s_2^2 = 5.0 \times 10^6$	$\sigma_I^2 + 2\sigma_M^2$
Between Channels within Motors	$f_1 = 1$	$s_1^2 = 1.0 \times 10^6$	σ_I^2

(b) ESTIMATES OF VARIANCE COMPONENTS

COMPONENT	ESTIMATE	COMMENT
σ_I^2	1×10^6	may be seriously underestimated.
σ_M^2	2×10^6	may be overestimated.
σ_{ML}^2	0	is consistently negligble.
σ_{BG}^2	0.2×10^6	

THE ANALYSIS OF WIND SPEED FREQUENCY DISTRIBUTIONS AND THEIR APPLICATION

Hans G. Baussus
Army Ballistic Missile Agency

1. INTRODUCTION. The above subject has been treated in some detail in an Aeroballistics Memorandum1). Since its publication on 21 June 1957 some further studies have been made which will be published about the end of November this year.

It is a well known fact that the wind, as a rather variable meteorological phenomenon, may be subject to a statistical investigation. As a measurable quantity in space and time it offers innumerable frequency distributions involving one and more variables and thus correlations.

2. Keeping some geographical region and some time interval constant, the wind may be regarded as a scalar which leads to one-dimensional frequency distributions for a certain altitude level. These distributions are generally skew with a mode smaller than the mean. In about 80o/o of all cases they obey a Pearsonian Curve of Type I which seems to fit the actual distributions better than Edgeworth's Series.

With respect to most applications where the wind is to be included in some functional expression, it has to be treated as a vector as in the case in analytical meteorology.

However, within a certain sufficiently small windrose sector the one-dimensional distribution holds. For vector statistics of winds see 2), 3).

It is clear that the more-dimensional distributions are generally non Gaussian ones although for most purposes these will yield enough information.

The statistical parameters to be derived in order to obtain the distribution desired are of course determined by samples. Wind measurements on a large scale have been carried out to altitudes up to 30 km, both the accuracy and sample size decreasing with increasing height. In most applications the vertical wind speed is neglected because of its comparatively low magnitudes and the fact that it can only be determined with a reliable accuracy by special devices.

Within some reference Cartesian coordinate system with the x,z plane

---

1. HANS G. BAUSSUS, The Analysis of Wind Speed Frequency Distribution and their Application, ABMA, Aeroballistics Memorandum No. 233, 1957 (secret)

2. C. S. DURST, Variation of Wind with Time and Distance, Geophysical Memoirs No. 93 (Great Britain), ASTIA Document No. AD 59531

3. HAROLD L. CRUTCHER, On the Standard Vector-Deviation Windrose, Journal of Meteorology, Volume 14, No. 1 (1957)

tangent to the earth (at a specific location) and the y axis perpendicular to this plane the normal distribution form would be for example

$$dH(x_1,x_2,z_3,z_4) = \frac{1}{(2\pi)^2\sqrt{\omega}} \exp\left[-\frac{1}{2\omega}\left(x_1'^2+x_2'^2+z_3'^2+z_4'^2 + 2\omega_{12}x_1'x_2'+2\omega_{13}x_1'z_3'+ 2\omega_{14}x_1'z_4' +2\omega_{23}x_2'z_3'+2\omega_{24}x_2'z_4' + 2\omega_{34}z_3'z_4'\right)\right] dx_1'dx_2'dz_3'dz_4' \qquad 4)$$

In this expression which is the standardized form of the normal distribution, the subscripts refer to altitude levels. Thus

$x_1' = \frac{x_1-\bar{x}_1}{\sigma_{x_1}}$ etc.,

$$\omega = \begin{vmatrix} 1 & \rho x_1x_2 & \rho x_1z_3 & \rho x_1z_4 \\ \rho x_1x_2 & 1 & \rho x_2z_3 & \rho x_2z_4 \\ \rho x_1z_3 & \rho x_2z_3 & 1 & \rho z_3z_4 \\ \rho x_1z_4 & \rho x_2z_4 & \rho z_3z_4 & 1 \end{vmatrix}$$

$\omega_{12} =$ etc.,

$$\rho_{12} = \frac{\frac{1}{n}\Sigma\,(x_1 - \bar{x}_1)(x_2 - \bar{x}_2)}{\sigma_{x_1}\,\sigma_{x_2}}$$

It appears to be possible to describe the main features of the horizontal wind frequency distribution by 10 variables.

Transformation formulas for a counterclockwise system are the following 5):

$\xi = x \cos\alpha = z \sin\alpha$,

4. cf. MAURICE G. KENDALL, The Advanced Theory of Statistics, Volume 1, London 1948.

5. HANS G. BAUSSUS, loc. cit., page 8

$$\zeta = -x \sin\alpha + z \cos\alpha,$$

$$\operatorname{var}\xi = \operatorname{var} x \cos^2\alpha + \operatorname{cov}(x,z) \sin 2\alpha + \operatorname{var} z \sin^2\alpha,$$

$$\operatorname{var}\zeta = \operatorname{var} x \sin^2\alpha - \operatorname{cov}(x,z) \sin 2\alpha + \operatorname{var} z \cos^2\alpha,$$

$$\operatorname{cov}(\xi,\zeta) = (\operatorname{var} z - \operatorname{var} x)\, 1/2 \sin 2\alpha + \operatorname{cov}(x,z) \cos 2\alpha$$

Without great difficulty, means and standard deviations of head and tail winds pertaining to a certain level and direction can be computed from the five statistical parameters necessary to establish the respective level distributions in the reference coordinate system.

The two-dimensional frequency distribution of the windshear 6), 7), which is the partial derivative of the horizontal wind velocity with respect to altitude can be derived from the respective wind frequency distribution. It is for example

$$\operatorname{var} S_{xm} = \left(\frac{\partial\sigma_{xm}}{\partial y}\right)^2 + \frac{1}{2}\sigma_{xm}^2 a_m \qquad \text{with } a_m \approx \frac{7 + \rho_{xmxm-2} - 8\rho_{xmxm-1}}{(y_m - y_{m-1})^2} > 0 \qquad 8)$$

3. Wind frequency as well as windshear distributions in their direct form are of interest to meteorologists, navigators, aerodynamicists and engineers working in the fields of structures and mechanics and guidance.

In a less direct form but nevertheless being of utmost importance, they enter the field of ballistics and are necessary for the calculation of firing table corrections thus improving the target hitting probability.

If again in a rectangular Cartesian coordinate system with the x, z plane tangent to the earth, the y axis perpendicular and positive upward and the x axis showing in the firing direction the differential equation with respect to x and valid for the dive phase of the ballistic missile assumed to have no terminal guidance writes as

$$\ddot{x} = -\frac{D}{m}\frac{1}{v^2}(\dot{x} - w_x)\sqrt{(\dot{x} - w_x)^2 + (\dot{y} - w_y)^2 + (\dot{z} - w_z)^2} \qquad 9)$$

---

6. NORMAN SISSENWINE, Windspeed Profile, Windshear, and Gusts for Design of Guidance Systems for Vertical Rising Air Vechicles, Air Force Surveys in Geophysics No. 57 (1954)

7. SIDNEY LEES, Study on Windshear Measurements, Quarterly Report under Corps Contract No. DA 36-039 SC-73204 (1957)

8. For correlation between levels see ARNOLD COURT, The Vertical Correlations of Wind Components, Scientific Report No. I Contract AF 19 (604)-2060, ASTIA Document No. 117182 (1957)

9. HANS G. BAUSSUS, loc.cit., page 11

or, after the root has been developed,

$$\ddot{x} + \frac{D}{m}\frac{1}{v}\dot{x} - \frac{D}{m}\left[\frac{v^2+\dot{x}^2}{v^3} w_x + \frac{\dot{x}\dot{y}}{v^3} w_y + \frac{\dot{x}\dot{z}}{v^3} w_z\right] = 0$$

By the methods of linear perturbations 10) the wind effect can be expressed as

$$\Delta x_w = \int_{to}^{tn}\left[\exp\left(-\int_{to}^{t}\frac{D}{m}\frac{1}{v}\,dt\right)\int_{t_0}^{t}\frac{D}{m}\left(\frac{v^2+\dot{x}^2}{v^3} w_x + \frac{\dot{x}\dot{y}}{v^3} w_y + \frac{\dot{x}\dot{z}}{v^3} w_z\right)\cdot\right.$$

$$\left.\exp\left(\int_{to}^{t}\frac{D}{m}\frac{1}{v}\,dt\right)dt\right]dt$$

or in an abbreviated symbolic form

$$\Delta x_w = \int_{to}^{t} F_1 W_x\,dt + \int_{to}^{t} F_2 W_y\,dt + \int_{to}^{t} F_3 W_z\,dt.$$

As the wind components are measured with respect to some Cartesian coordinate system in the dive phase region, a coordinate transformation must be applied. In order to simplify and generalize the whole method, the final corrections should be expressed in the target coordinate system X,Y,Z with the X axis showing in East and the Z axis showing in South direction. The wind coordinates may be once for all determined pertaining to this system. With λ_2,Φ_2 as the geodedic coordinates of the missile launching point, λ_2,Φ_2 those of the target, the procedure is the following (λ counted from East to West, $\Delta\lambda = \lambda_1 - \lambda_2$):

1) $\sin\alpha = \dfrac{\sin\Delta\lambda \quad \cos\Phi_2}{\sqrt{1-(\sin\Phi_1 \sin\Phi_2 + \cos\Phi_1 \cos\Phi_2 \cos\Delta\lambda)^2}}$

2) $\tan\sigma = \frac{z}{x} \rightarrow \sigma$ which is a small angle being a function of latitude, azimuth and distance.

3) $\delta = 90^\circ - \alpha + \sigma$,

4) $a_1 = \cos\delta \cos\Delta\lambda + \sin\delta \sin\Phi_1 \sin\Delta\lambda$

$a_3 = \cos\delta \sin\Phi_2 \sin\Delta\lambda - \sin\delta(\cos\Phi_1 \cos\Phi_2 + \sin\Phi_1 \sin\Phi_2 \cdot$

$\cos\Delta\lambda)$,

10.see, e.g., JOHN W. GREEN, Exterior Ballistics, in Edwin F. Beckenbach, Modern Mathematics for the Engineer, New York 1956

$$b_1 = -\cos\phi_1 \sin\Delta\lambda,$$

$$b_3 = \cos\phi_1 \sin\phi_2 \cos\Delta\lambda - \sin\phi_1 \cos\phi_2,$$

$$c_1 = \sin\delta\cos\Delta\lambda - \cos\delta\sin\phi_1 \sin\Delta\lambda,$$

$$c_3 = \sin\delta\sin\phi_2 \sin\Delta\lambda + \cos\delta(\cos\phi_1 \cos\phi_2 + \sin\phi_1 \sin\phi_2 \cos\Delta\lambda).$$

5) $$\Delta x_3 = a_1 \int_{to}^{tn} F_1 W_x dt + a_3 \int_{to}^{tn} F_1 W_z dt$$

$$\Delta y_3 = b_1 \int_{to}^{tn} F_4 W_x dt + b_3 \int_{to}^{tn} F_4 W_z dt,$$

$$\Delta z_3 = c_1 \int_{to}^{tn} F_7 W_x dt + c_3 \int_{to}^{tn} F_7 W_z dt,$$

6) $$-\Delta X = (\cos\delta \cos\Delta\lambda + \sin\delta\sin\phi_1 \sin\Delta\lambda)\Delta x_3 - \cos\phi_1 \sin\Delta\lambda\, \Delta y_3 + (\sin\delta\cos\Delta\lambda - \cos\delta\sin\phi_1 \sin\Delta\lambda)\, \Delta z_3$$

$$-\Delta Y = \left[\cos\delta\cos\phi_2 \sin\Delta\lambda - \sin\delta(\cos\phi_2 \sin\phi_1 \cos\Delta\lambda - \sin\phi_2 \cos\phi_1)\right]\Delta x_3 + (\sin\phi_1 \sin\phi_2 + \cos\phi_1 \cos\phi_2 \cos\Delta\lambda)\, \Delta y_3 + \left[\sin\delta \cdot \cos\phi_2 \sin\Delta\lambda + \cos\delta(\cos\phi_2 \sin\phi_1 \cos\Delta\lambda - \sin\phi_2 \cos\phi_1)\right]\Delta z_3$$

$$-\Delta Z = \left[\cos\delta\sin\phi_2 \sin\Delta\lambda - \sin\delta(\cos\phi_1 \cos\phi_2 + \sin\phi_1 \sin\phi_2 \cdot \cos\Delta\lambda)\right] \Delta x_3 + (\sin\phi_2 \cos\phi_1 \cos\Delta\lambda - \cos\phi_2 \sin\phi_1)\, \Delta y_3 + \left[\sin\delta\sin\phi_2 \sin\Delta\lambda + \cos\delta(\cos\phi_1 \cos\phi_2 + \sin\phi_1 \sin\phi_2 \cdot \cos\Delta\lambda)\right] \Delta z_3.$$

Terms of minor influence have been omitted here. F_4 and F_7 are the analog functions to F_1. It may be noted that for longer distances a correction term ΔY enters due to the curvature of the earth.

true, the hitting accuracy can be increased by a factor $k \approx 3$ for the probability levels 0.5 to 0.9 14).

The component σ_w^2 (moment about the mean) of var X_2 could be considerably reduced if the wind direction (s) were known, which can certainly be derived or extrapolated from synoptic weather charts. Even lower winds, in nearly all cases, will give the wind direction quite accurately 15). In this case the sector or one-dimensional wind distribution would replace the two-dimensional one. It should be emphasized at this point that the possible application of such a method which may be combined with some statement as "weak," "average," "strong," would yield results needing no more refinement, as it is not possible to remove the other components of the total variance.

There exist several methods based on multiple regression which may be used to forecast pressure and wind fields 16),17),18). In a number of cases they can be utilized and may be of considerable value. However, it is the integration procedure which enters in addition, requiring more valuable time. Certain relationships exists between 2) and 18). If C. E. Buell's estimation of winds based on the departure of the height of a constant pressure surface from its mean value proves to be satisfactory for levels other than the 500 mb one and especially for the Russian continent, it may become a very convenient tool bringing the wind effected variance down by about 40 per cent.

Mean patterns of winds in partly different forms have been provided by different authors 19),20),21). whereas sector distribution parameters

---

14. HANS G. BAUSSUS, Loc. cit., pages 13, 14, 47

15. See: The Jet Stream, The Chief of the Bureau of Aeronautics, 1953, ASTIA Document No. 36014, page 57.

16. Short Range and Extended Forecasting by Statistical Methods, Headquarters Air Weather Service, Washington, D. C., 1948.

17. Practical Methods of Weather Analysis and Prognosis, Office of the Chief of Naval Operations, 1952, ASTIA Document No. AD 35603

18. C. F. BUELL, The Correlation between Wind and Height on an Isobaric Surface, The Kaman Aircraft Corporation, Albuquerque, New Mexico, 1957.

19. RICHARD SCHERHAG, Neue Methoden der Wetteranalyse und Wetterprognose, Berlin 1948

20. BROOKS et. al., Upper Winds over the World, Geophysical Memoirs No. 85 (Great Britain) 1950

21. A. F. JENKINSON, The Average Vector Wind Distribution of the Upper Air in Temperate and Tropical Latitudes, a paper of the Meteorological Research Committee (London), ASTIA Document No. AD 37 923

are not available yet. As a matter of fact, their determinations would require a larger statistical population. Assuming normality of the two-dimensional distribution it would of course be possible to derive the sector distributions analytically.

4. SUMMARY AND CONCLUSIONS. Wind frequency distributions, several classes of which can be derived or investigated, seem to be a source of interest today not only to the meteorologist. The distribution of the mean wind over the world and its utilization in the dive phase analysis of large range ballistic missiles contributes to the improving of target hitting probability. Of particular interest are the levels between 4 and 24 km altitude, however, for certain warheads information up to 35 km is essential. ICBM's might even require some wind statistics beyond this level. Sector distribution parameters and some additional method which increases the accuracy and does not absorb much time, should be supplied at the earliest possible date. As for purely statistical methods depending on no additional or actual information, firing correction tables can and should be prepared in time including both mean wind and mean density material.

REFERENCES

For several additional or other references not mentioned in the footnotes see list in Hans G. Baussus, loc. cit.,

22. T. OZAWA and K. TOMATSU, Application of Momentum Transport Theory for 5-Day Mean Chart, Meteorological Research Institute, 1956

23. BERT BOLIN, Studies of the General Circulation of the Atmosphere, University of Stockholm (Sweden)

24. SVERRE PETTERSSEN, Weather Analysis and Forecasting, Volume I, II, New York 1956

25. T. BERGERON et. al., Dynamic Meteorology and Weather Forecasting, Washington, D. C., 1957

26. H. KOSCHMIEDER, Dynamische Meteorologie, Band 2, Leipzig 1951

EXPERIMENTAL INVESTIGATION OF THE MOTION OF A LIQUID IN A DECELERATED GUIDED MISSILE CONTAINER

E. A. Hellebrand
Army Ballistic Missile Agency

INTRODUCTION. When the thrust force of a missile is terminated, the liquid remaining in its tanks, due to drag forces decelerating the missile body, moves to the front and impinges on the bulkhead. A missile laid out for a certain range will have a considerable amount of liquids left in the containers, when fired over a shorter distance. Since in this case, cut-off occurs at low altitude, the drag is also quite substantial. In extreme cases, several thousand pounds of liquid impinge on the bulkhead, but the mode of impact as well as the velocity and pressure distribution were unknown. The problem was to find out whether the bulkhead designed for static pressure would stand the impact loads also. About 15 years ago, a series of experiments were run in the structures test laboratory of the German Army Rocket Center in Peenemuende. Test results indicated a liquid movement comparable to free stream conditions encountered on the buckets of a Pelton Turbine Wheel. However, the model container was very small, film coverage was not quite adequate and no force or pressure measurements were taken. Consequently, design criteria could not be derived from the rather scarce test data.

The Structures & Mechanics Laboratory, Development Operations Division, ABMA, therefore asked the Southwest Research Institute in San Antonio, Texas to conduct experimental research to clarify the behavior of the liquid in missile containers with various amounts of liquid and at different deceleration levels.

This paper attempts to describe the more important considerations that determined the general experimental set-up, including model analysis, test apparatus, instrumentation, performance, and test evaluation and also to indicate ways of analytical approximate prediction of the forces on bulkheads and walls of a missile container under these conditions.

FLIGHT CONDITIONS. The missile attitude at thrust termination is indicated in Figure 1.* Thrust decay causes a rather abrupt change from acceleration to deceleration. (Deceleration shown as shaded area). The deceleration due to drag will seldom reach more than 0.2 g, but in extreme cases up to 0.6 g can be expected. This force acts on the missile body only, giving rise to a relative forward motion of the remaining liquid until the bulkhead partially stops and reverses the motion.

TEST SET-UP. In order to stimulate flight conditions on the ground, the test frame shown in Figure 2 was developed and built by Southwest Research Institute. The outer tubular frame serves as pressure accumulator. The upper piston is forced down by the pressure p_1 entering through Valve No.1. The ensuing down stroke of the piston stimulates the missile condition until the liquid has impinged on the forward bulkhead. Now the piston passes Valve

*Figures are at the end of the article.

No. 2, which vents the accelerating pressure. Immediately thereafter the piston pushes against an air cushion for smooth braking, which forces a floating piston further down against the action of pressure p_2, until all energy is absorbed. Valve No. 3 then reduces p_2 in such a way that rebound is minimized. The model tank rails rigidly connected to the piston rod, is guided absolutely smoothly along two rails. Chatter and vibration were practically absent due to the excellent quality of the structure. The test frame can be tilted to simulate different missile attitudes. Figure 3 indicates the camera position with respect to the inclined test frame, the model structure, screen and lights.

MODEL ANALYSIS. To insure proper dynamic simulation, Southwest Research Institute conducted a model analysis. With seven important parameters given:

d	=	Tank Diameter
a	=	Acceleration
ρ	=	Density of Liquid
μ	=	Viscosity of Liquid
σ	=	Surface Density of Liquid
p	=	Pressure
t	=	Time

and three possible combinations of exponents used up to satisfy dimensional agreement in lengths, mass, and time, there are four dimensionless groups left (π_1 through π_4). π_3 and π_4 can also be derived from the law of capillarity and the compatibility of Reynold's numbers respectively. See Figure 4.

From π_3 we obtain the diameter ratio of model versus prototype tanks. With d found, the pressure ratio is obtained from π_1, the acceleration from π_4 and time ratio from π_2. Results are given on the bottom line of Figure 5.

With kerosene as the prototype liquid, water as a model liquid would require unduly high accelerations with a corresponding reduction to extremely small model tanks. Carbon tetrachloride allows for a larger model with reduced force requirements and a reasonably long model time. The advantages of carbon tetrachloride are its high specific density, low viscosity and surface density. All measuring tests were run with carbon tetrachloride. The properites of water, carbon tetrachloride and kerosene are given on Figure 5, together with diameter, acceleration, time and pressure resulting from the application of these liquids.

FLOW PATTERN. If the original liquid surface is perpendicular to the drag forces, a central column of liquid rises from the surface with an annular void appearing at its root. This is due to wall friction, surface tension, and viscosity holding back the liquid close to the tank walls. See Figure 6. Quantitative evaluation of the volume of the liquid column and the void at

the bottom at different times during the downward stroke, seems to indicate that the density of the liquid in the column is greatly reduced to 1/2 - 1/3 of the original value. This agrees quite well with impact pressures measured on the forward bulkhead which are generally smaller than expected from the density of the liquid at rest. A semi-empirical formula utilizing the dynamic pressure notations: $p = C\rho_1 \mu^2/2$ with $\mu = \sqrt{2ah}$ results in

$$p = C \frac{\rho_1}{\rho_0} \rho_0 ah = C \frac{\rho_1}{\rho_0} \gamma_0 nh$$

where:

C = drag or form factor (impact pressure coefficient)

p = average impact pressure, psi

ρ = density of liquid, lb sec^2/in^4.

γ = specific weight of liquid lb/in^3

a = acceleration, in/sec^2

h = average distance from original surface to impact area, in.

If the original surface is not perpendicular to the drag action, the picture changes rapidly. A quite solid wave starts running up at the side toward which the original liquid surface leaned and impinges heavily on the head. The surface does not break up and the density is not reduced. The different flow conditions are expressed by the form Factor C, which varies with the shape of the bulkhead and the location of the first contact. It is highest with the inverted cone head and tilted stroke due to a wedging action of the liquid at the relative narrow annular space between tank wall and cone. It is lowest if on the vertical stroke an inverted cone is used which deflects the liquid relatively smoothly towards the outer tank walls. The following table 1 gives the average magnitude of C and ρ_1/ρ_0 fitted to a great number of tests versus bulkhead geometry and angle of tank inclination α.

TEST RESULTS. Figures 7 and 8 show test results versus numerical values of equation 1, which quite clearly indicate the linear relationship of pressure and acceleration, if the other parameters are kept constant. Test results on Figure 8 show a larger scatter due to a strong local turbulence caused by the inverted cone. The straight lines in both figures represent equation 1.

Figure 9 illustrates a typical oscillograph record of the pressure cells and accelerometer readings versus time on the spherical head with $\alpha = 50°$ Cell #1 indicates pressure build-up first, closely followed by the others. Equation #1 applied to this special case with $a_{avg} = 35g$, $h_{avg} = 9$ inches, $c = 1$ and $\rho_1/\rho_0 = 1$ and a specific density of carbon tetrachloride of 0.058 pounds per cubic inch gives a model pressure of 18.3 psi. The corresponding prototype conditions are $a_{avg} = 0.4$ g, h = 110", and the pressure p = 1.4 psi. Measured model pressure in this case were 19.3 psi taken as the average of all 4 cells. Equation 1 thus underestimates the pressure by about 6%, which is acceptable considering the complexity of the flow. The model flow velocity shortly before impact is 490 in/sec. The corresponding prototype velocity is 180 in/sec and the Reynolds Number in both cases is 2.8×10^6,

with the tank diameter as reference, far above the critical value.

In order to better simulate missile tank conditions, a number of tests were made with plastic model rings glued to the tank walls. The rings had a depth of 1/8 in to coincide with the prototype ratio of tank diameter vs ring depth and had a square cross section. The presence of the rings created additional friction and turbulence in the moving liquid and reduced the average impact pressure by 20 to 30 percent.

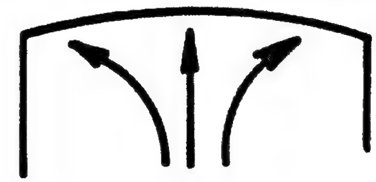
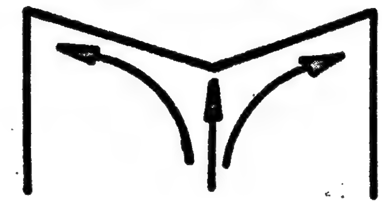
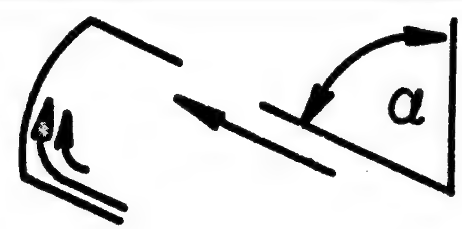
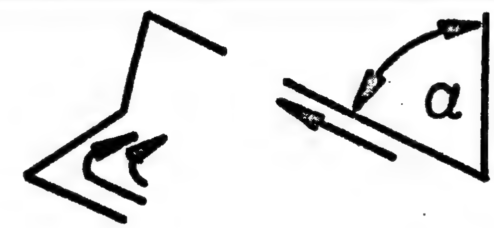
$\alpha = 0^{\circ}$		
C_D \ ρ_1/ρ_0	1.0 \ 0.5	0.5 \ 0.5
$\alpha = 25^{\circ}$		
C_D \ ρ_1/ρ_0	1.0 \ 1.0	2.0 \ 1.0

TABLE I

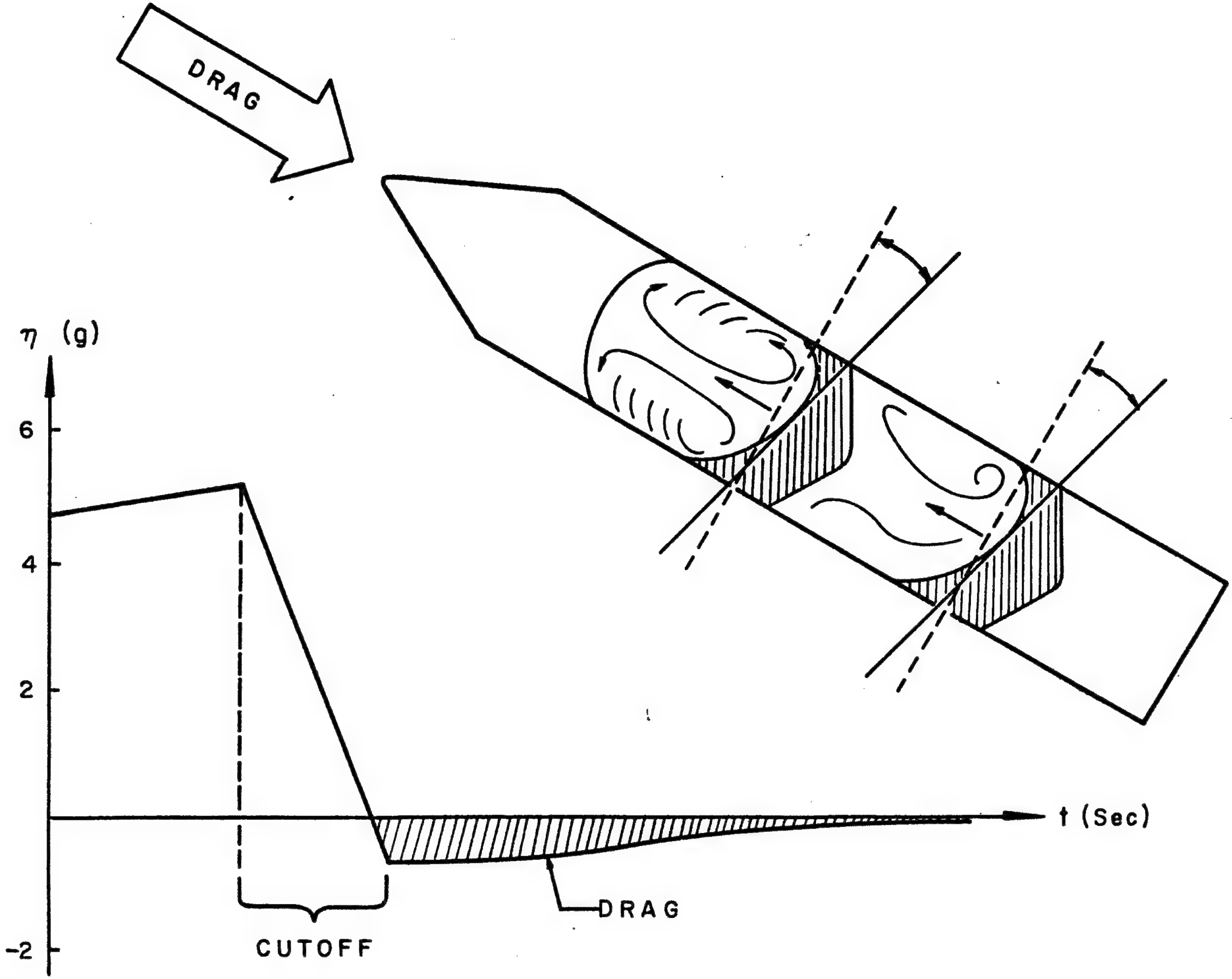
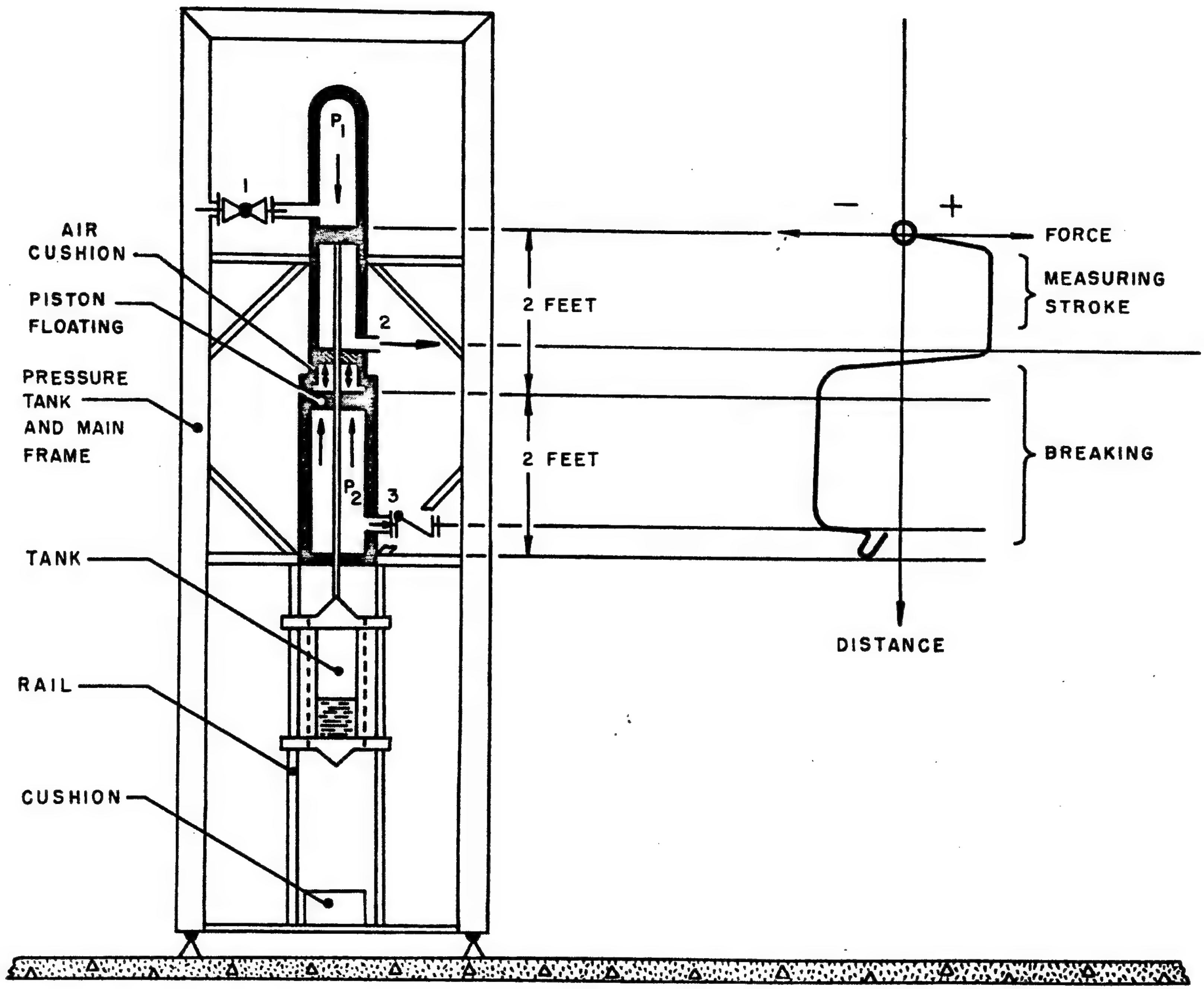


FIG. 1 MISSILE AT THRUST TERMINATION


AIR CUSHION
PISTON FLOATING
PRESSURE TANK AND MAIN FRAME
TANK
RAIL
CUSHION
P1
P2
1
2
3
2 FEET
2 FEET
FORCE
+
−
MEASURING STROKE
BREAKING
DISTANCE

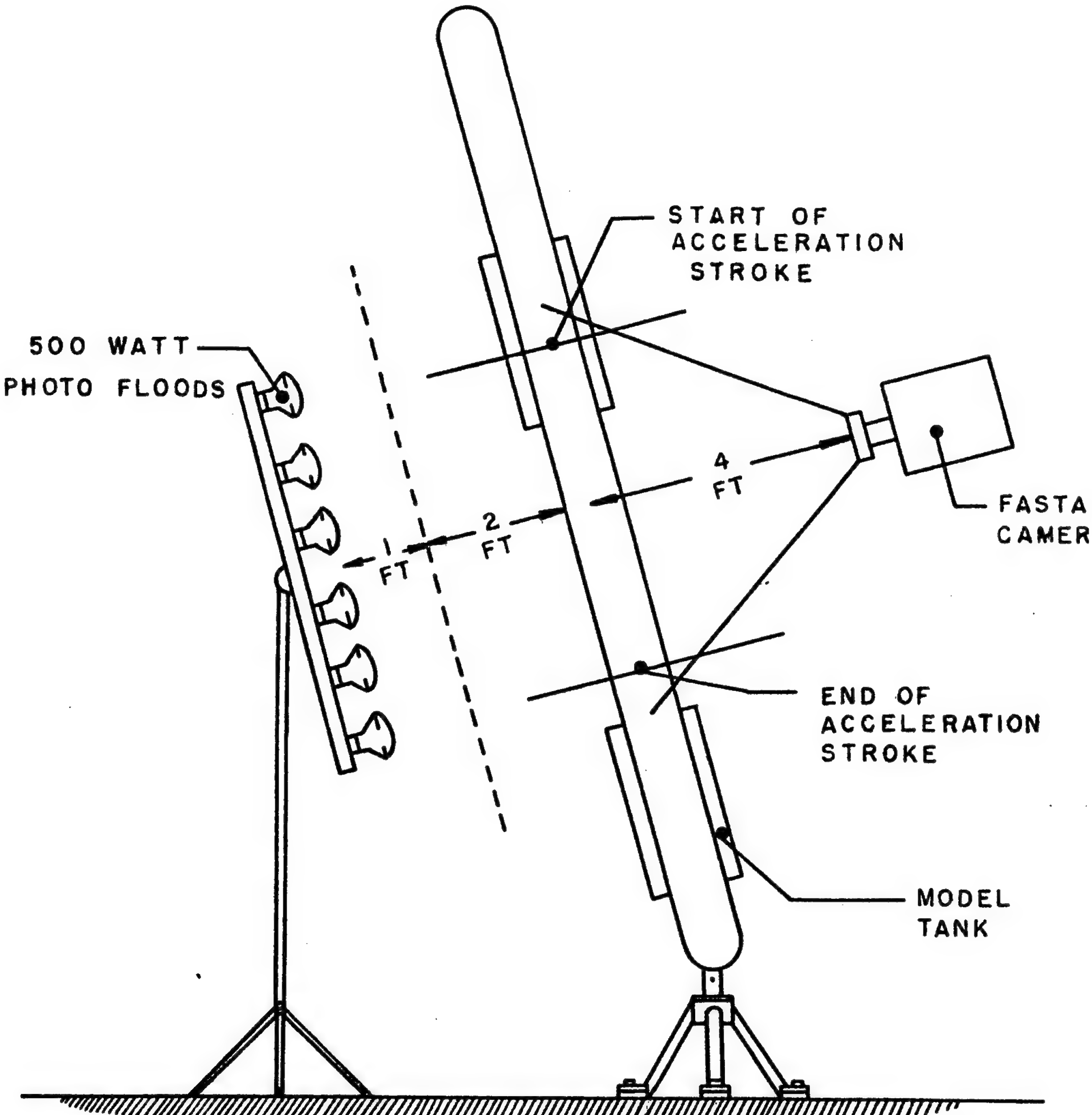


FIG. 3 - CAMERA POSITION

4: Model Analysis

$\pi = d^a \rho^b \mu^c a^d \sigma^e p^f t^g$ dimensionless group

$= L^a (ML^{-3})^b (ML^{-1} T^{-1})^c (LT^{-2})^d (MT^{-2})^e (ML^{-1} T^{-2})^f (T)^g$

$a - 3b - c + d - f = 0$ Condition on L

$b + c + e + f = 0$ M

$c + 2d + 2e + 2f - g = 0$ T

$\pi_1 = pd^2\rho/\mu^2$ pressure $f = 1$ $d, e, g = 0$

$\pi_2 = \mu at/\sigma$ time $g = 1$

$\pi_3 = \sigma d \rho/\mu^2$ surface tension $e = 1$ directly from law of capillarity

$\pi_4 = ad^3 \rho^2/\mu^2$ acceleration $d = 1$ directly from Reynold's Number

5. Model Analysis (continued)

Proper modeling demands: $\pi_{im} = \pi_{ip}$ or: $p_R d_R^2 \rho_R / \mu_R^2 = 1$ a. s. o.

MODEL		PROTOTYPE
Carbon Tetrachloride	Water	Kerosene
$\rho = 1.60$	1	0.83 gm/cm^3
$\mu = 0.96$	1	2.50 centipoise
$\sigma = 27$	72	27.5 dynes/cm
d = 5.5"	3.5"	70"
a = 8 - 48 g	89 - 540 g	0.1 - 0.6 g
t = 0.031 sec	0.0073 sec	~1 sec
p = 1 - 30 psi	5 - 160 psi	0.1 - 3 psi

$d_R = 0.08$ $a_R = 82.5$ $t_R = 0.031$ $p_R = 12.4$ — for Carbon Tetrachl. vs. Kerosene

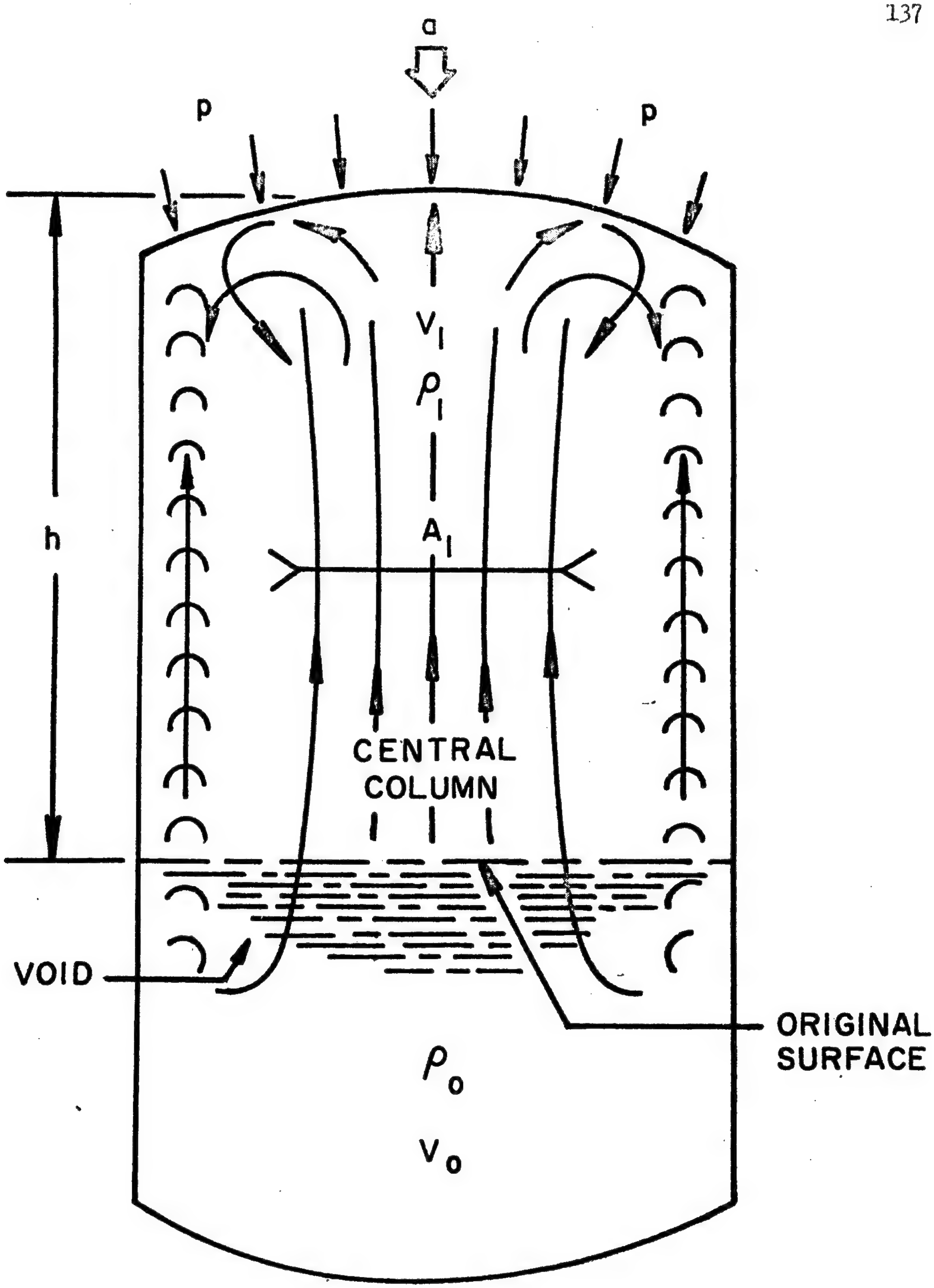


FIG. 6 FLOW ANALYSIS

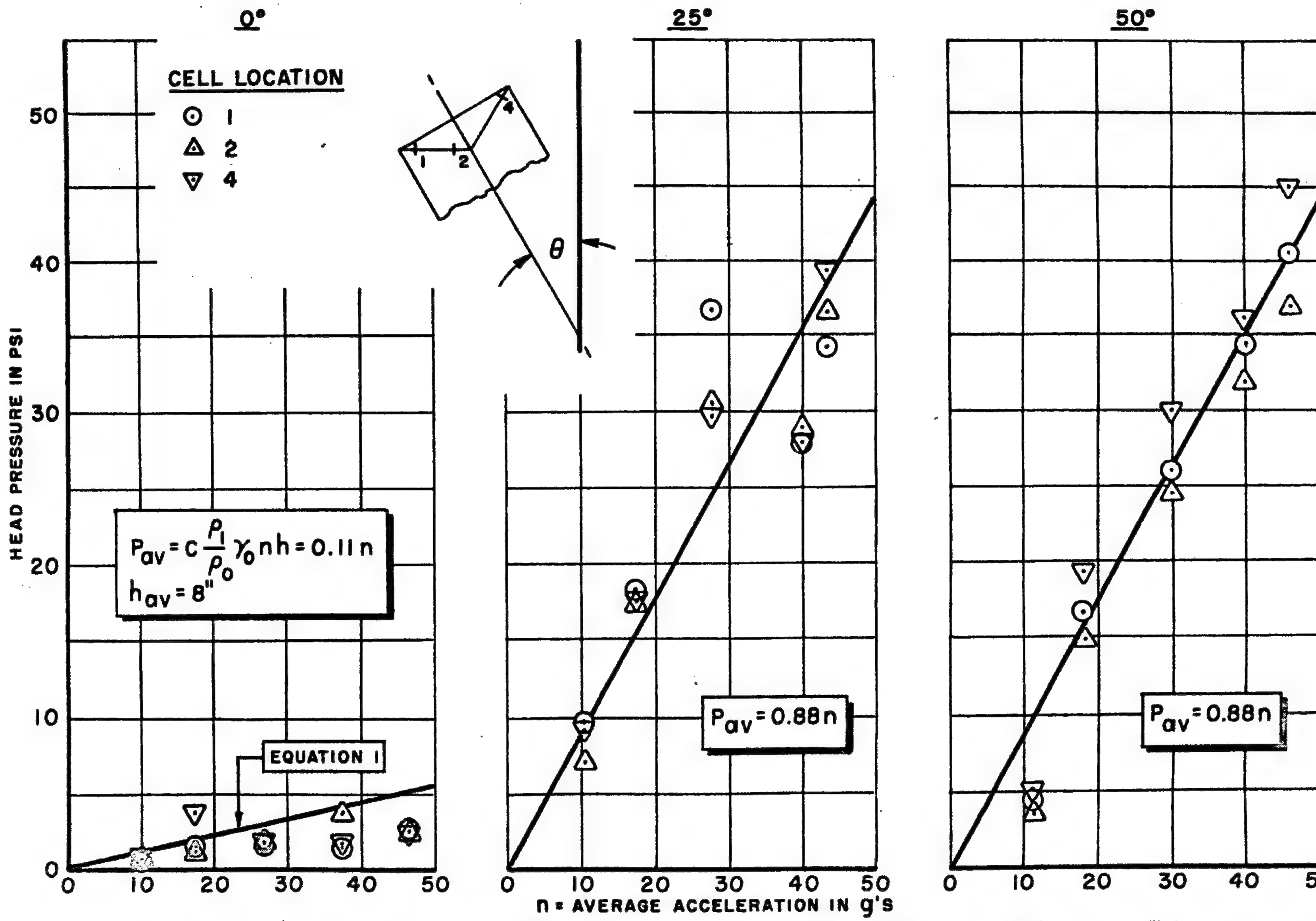


FIG. 7 HEAD PRESSURE VS ACCELERATION. CONICAL HEAD. 1/4 FULL

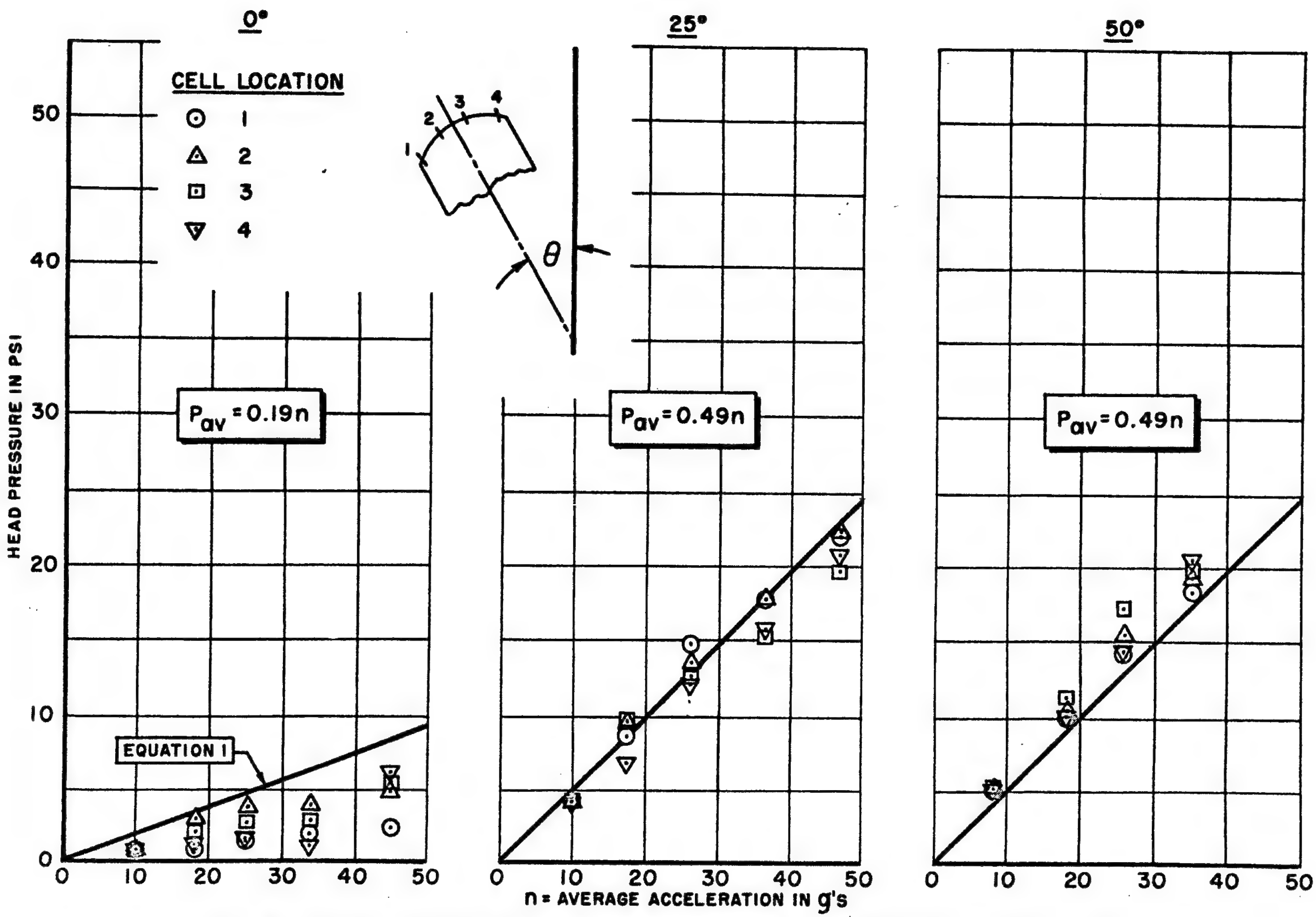


FIG. 8 HEAD PRESSURE VS ACCELERATION, SPHERICAL HEAD, 1/4 FULL

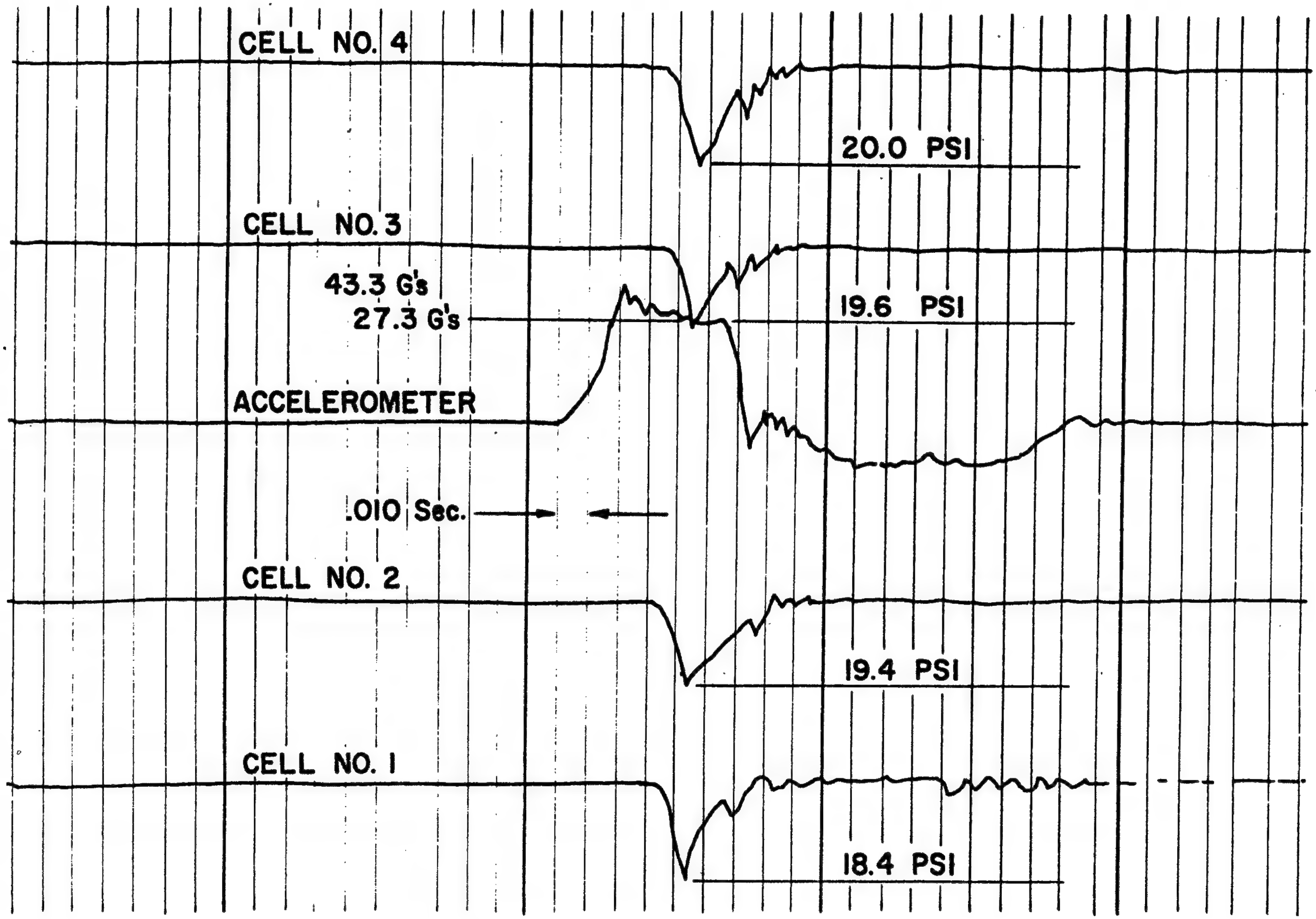


FIG 9. TYPICAL OSCILLOGRAPH RECORD

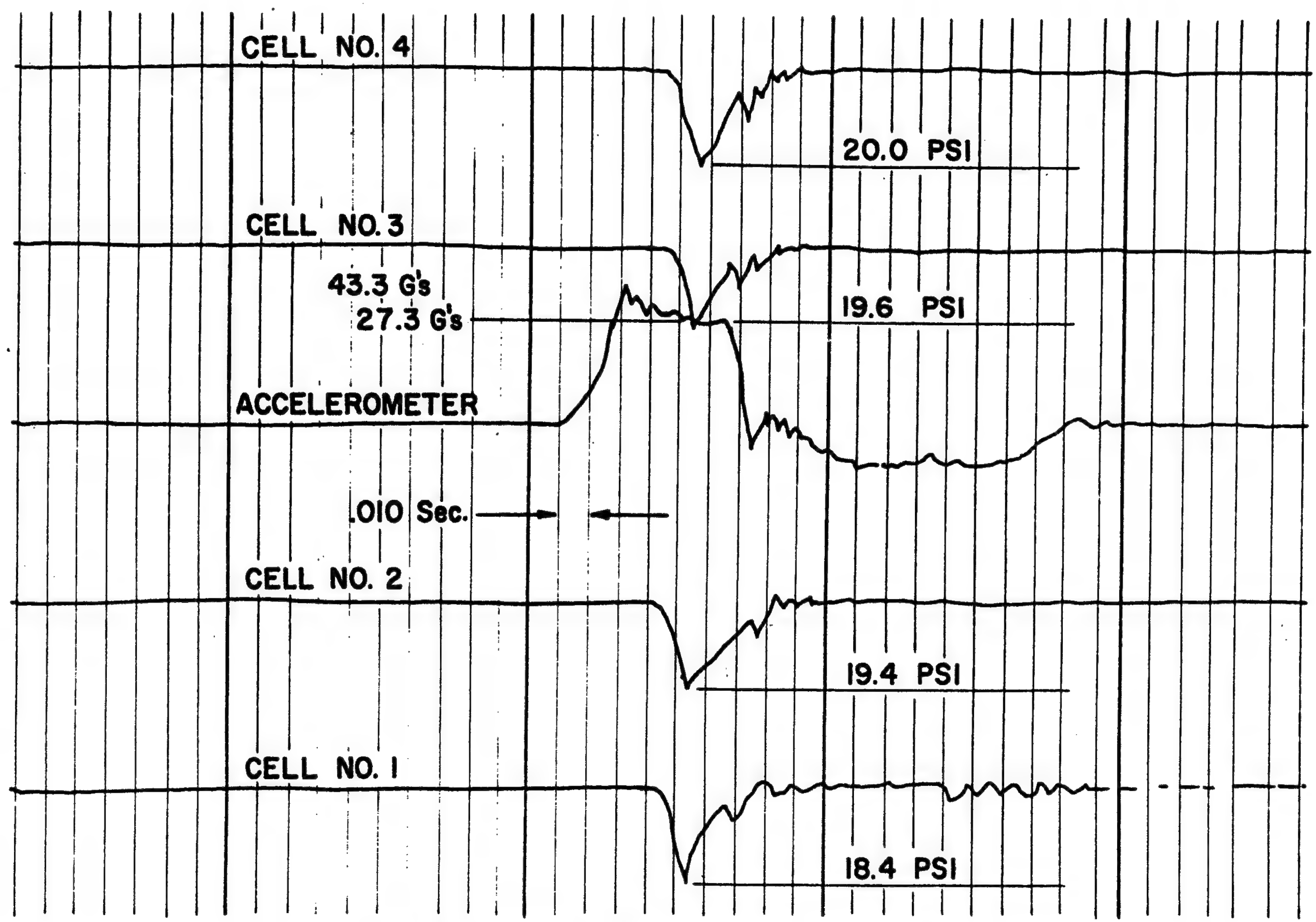


FIG 9. TYPICAL OSCILLOGRAPH RECORD

AN EXAMPLE OF AUTOMATION WITH ASSOCIATED STATISTICAL PROBLEMS

E. L. Cox and W. D. Foster
Program Research Branch, Assessment Division
Fort Detrick, Maryland

The term automation is now generally applied in a technical sense to that kind of machine which performs a sequence of operations without human guidance; that is, there is a pre-set "program" which after the machine is set in operation continues a further operation at the completion of the preceding until all the indicated operations have been performed. We wish to use this term in a somewhat more general sense to include instruments which may or may not perform a sequence of operations but do substitute a machine for a sequence of human operations resulting in quantitative data expressed as counts, measurements, dial recordings, et cetera.

The main body of this discussion will be centered on the operation of the DAC, an abbreviation which represents the DuMont Automatic Counter. A fundamental process in bacteriological laboratories is the preparation of plates containing colonies. On these plates, small dishes containing about ½ centimeter thickness of agar, a dilute bacterial broth is poured. The single bacterial cells multiply under incubation to produce clumps, known as colonies, which are visible to the unaided eye. The operations on these plates which are to be superseded by the automatic machine are (a) observing the plates under magnification, (b) counting visually the number of colonies appearing on the plate, (c) recording the counts on a hand operated device, (d) entering these counts on tabular paper. From the statisticsl point of view there are questions of error associated with each of these operations. These will be discussed later.

When a plate is placed in the DAC, a light beam passes through the plate in a sequence of scans so that the beam starting at one side of the plate in its scanning action completely "observes" the whole surface of the plate. If the light beam is interrupted by a colony or any other opague object in its path, this interruption activates an electronic mechanism which registers a count. The machine has a "memory" so that if an interruption occurs at the same place on adjacent scans, an additional count is not made. After the scan has passed completely across the plate, the machine further operates to transfer the accumulated count for that plate to a printing mechnaism.

In considering the possible sources of errors by technicians in the counting procedure, it has been observed that problems associated with the observation of the plates, the recording of counts on the counter and the entering of the number on a table are small in magnitude. The major error in this operation comes from the technicians failure to count certain colonies or in counting colonies more than once. The resultant recorded counts are found to be normally distributed about a true count. (With certain technicians a definite bias seems to be discerned.) In general, also, the variability about the true count seems to be a function of this true count level; that is, the larger the mean number of colonies on a plate the larger the measure of variability from the count observed by the technician will be. The DAC being a machine does not observe plates in the sense that a human operator does. The machine cannot discriminate between touching or overlapping colonies nor can it record colonies at the very edge of the plate

which is outside the bounds of the scanning mechanism. Because touching and overlapping increase with larger numbers of colonies on the plate, the mean number of colonies reported for any plate by the machine is biased downward as a function of the number of colonies present. Moreover, careful study of the operation of the machine has discerned that while a count reported by the machine for any one plate may be well reproduced by further recordings on the same plate, recordings on different plates showing the same number of colonies will have a variability from the mean somewhat greater than that provided by repeated technician counts on the same plate. Major problems requiring statistical attention in counting plates by machine, then, are the development of an expression to relate mean machine counts with true counts and the estimation of the variability of individual counts as they depart from the true values.

To develop an expression which would permit machine counts to be related to the true counts on the plates, it was necessary to generate pairs of values expressing these two count properties. If the relationship between the pairs of values had been indicated as 1 to 1 with little or no variation from this property, a rather simple relationship would have been evidenced. Moreover, even if considerable variations had been present and the mean relationship had been 1 to 1, this relationship would probably have been used as a basis for comparison. However, as both departure from linearity and change in variation was indicated throughout the range studied, it was necessary to study the problems with more care. It was found that transformation of each member of the pairs of observations to its corresponding logarithm provided sets of values which when plotted seemed to conform reasonably well to a linear model. Moreover, the variation in the logarithms was nearly constant over the whole range of observation. Regression techniques were considered appropriate for determining a descriptive equation relating machine counts to true counts. From this equation it is possible to develop from any given machine count an estimate of the number of colonies actually appearing on a plate and to compute bounds on this estimate which indicate a percentage error that may be present. As the variance associated with estimates made from the machine counts were not noticeably greater than that occurring from the normal practice of counts by technicians of the same material, it was argued that counts produced by the machine could be as accurate as those obtained by the previously used technique. The advantage then lay with the speed of the machine. From this argument came the recommendation that the machine be accepted and put into use.

In general, a machine or instrument will have a different criterion for discrimination from that provided by the operation that it is to supersede. The important problem in investigation is to discover the nature of such discrimination and, if possible, to relate it to properties better understood and more easily described. For example, another instrument which we have examined is a "chart recorder." There is a number of devices which provide information by drawing a continuous trace on a moving roll of paper. We use such an instrument, the Esterline Angus Recorder, for making a record of windspeeds and directions over a measured period of time. While these traces give a record of the variability measured and its changes with time, the conversion of this record to useful digital information is not easily obtained. A device which reads a measurement at definite distances along the chart (and hence at definite times) provides well defined information at those time instances but loses some information about the variability

of the changes with time which are so graphically presented in the moving trace which has been drawn on the chart roll. The relationship between the information given on the chart roll and by the chart reader comprise a calibration problem complicated by a sampling procedure.

With other instrumentation and machine development of data there are similar problems. With each there are statistical problems. In general, however, the problems can be resolved into those of finding a functional relationship between pairs of values and the assessment of this relationship in terms of the magnitude of the variance associated with the measurements observed.

EXPERIMENTAL DESIGN TO STUDY THE EFFECT OF BALLOON SIZE ON WIND RESPONSE

Raymond Bellucci
Meteorological Division
U. S. Army Signal Engineering Laboratories

INTRODUCTION. Measurements of wind speed and direction above the earth's surface are made by tracking the path of a freely ascending balloon or balloon-borne equipment. Observations of the balloon's position and height at the beginning and end of a time interval give the necessary data for computing the mean wind speed and direction through the layer.

The pilot balloon, a free balloon whose movements can be observed by a theodolite, was first used in 1909 for upper-air wind measurements. This small, spherical balloon weighs 100 grams or less, its size depending on the ultimate height of the observations, and is approximately three feet in diameter when inflated at the ground. Before World War I, this balloon had been widely adopted; during the war, the need for accurate observation of upper winds was urgent in both the artillery and aviation services, and use of the pilot balloon became firmly established; today, pilot balloon data are used by the U. S. Weather Bureau and the USAF for obtaining winds to 40,000 feet.

Interest in obtaining accurate meteorological data above 100,000 feet resulted in development of larger balloons. The first ones developed for this purpose at the U. S. Army Signal Engineering Laboratories were manufactured by Molded Latex Products, Inc. The balloons reached an altitude of 120,000 to 140,000 feet during the daytime. They carried a payload of about 2,000 grams, weighed approximately 10,000 grams each, had an overall length of around 20 feet, and required nearly 700 feet^3 of gas for inflation. To attain an altitude above 100,000 feet, the balloons were only partially inflated at the ground and became fully extended at about 30,000 feet. Their rate of climb was 800 feet min.$^{-1}$ to 30,000 feet and 1,100 feet min.$^{-1}$ to burst.

The 100-gram balloon maintains its spherical shape throughout flight; whereas, the large 10,000-gram balloon is distorted in flight until it becomes fully extended at approximately 30,000 feet, whereupon it becomes spherical. The question arose as to whether or not the difference in size and shape of the balloons had any effect on their wind response; that is, are the wind speeds that are determined by tracking these balloons equivalent? An experiment was designed to answer this question.

For this study two balloon types, the 350-gram and the 10,000-gram, were selected because they represent the extremes in size, shape, and texture. Therefore, any differences that exist will be magnified.

PROPOSED EXPERIMENT

BACKGROUND. Assume that the wind in a given space and time interval is composed of n extremely small parcels of air whose velocities are known. Then the true mean velocity of the mass for this interval is the vector sum of the velocities of the parcels divided by n. The differences in magnitude and direction of the velocities of these parcels from the true

mean are due to the natural variability of the wind. The observed mean wind measured by tracking a balloon for a given time interval is an approximation to the true mean wind. This approximation is due to the balloon's imperfect responsiveness to the wind and to the error in the tracking system. Therefore, the observed variance from the true mean in a given time interval is equal to the sum of the variances due to natural variability, balloon response to the wind, and tracking system error. Symbolically, this can be written as follows:

$$S_o^2 = S_r^2 + S_v^2 + S_T^2$$

Where S_o^2 is the observed variance,

S_r^2 is the variance due to balloon response,

S_v^2 is the variance due to natural variability, and

S_T^2 is the variance due to the tracking system.

Let $\vec{V}_a$ and $\vec{V}_b$ be the wind velocities obtained by tracking balloons of different types. Then if two balloons of the same type are released simultaneously and vector velocity differences are computed for each 1,000-foot level, these differences $(\vec{V}_{a_1} - \vec{V}_{a_2})$ will not reflect the balloon responsiveness to the wind, but will represent the combined effects of tracking errors and natural variability of the wind.

The simultaneous release of two balloons of different types will give vector velocity differences $(\vec{V}_a - \vec{V}_b)$, based on the aforementioned causes plus the difference in balloon response to the wind. The only change then is that the responsiveness of two different types of balloons is considered in the latter case. Therefore any difference in the mean value of $|\Delta V_{a-b}|$ and $|\Delta V_{a_1-a_2}|$ should be due to a difference in the responsiveness of the two balloon types.

The null hypothesis to be tested is that the difference between the means is zero, i.e. $\overline{|\Delta V_{a-b}|} - \overline{|\Delta V_{a_1-a_2}|} = 0$. The 95-percent confidence level will be used to determine whether or not the computed difference is significant.

If the difference is significant at the 95-percent level, it can be concluded that the results are not consistent with the hypothesis of equal means. Therefore, the difference in the mean is caused by differences in balloon types. However, if the difference between the means is not significant, no definite conclusions can be drawn except it appears that different balloon types do not significantly affect the balloon response to the wind. It also indicates that the experiment should be extended or that a different experimental approach to the problem is needed.

DESCRIPTION. The experiment will consist of simultaneously releasing three balloons--one 10,000-gram and two 350-gram. Four radio-direction-finding sets (GMD-1) will be used to track the balloons in flight. One balloon will be tracked by sets one and two; the other two balloons by sets three and four. The fourth or control set will be paired with one of the other sets for a particular flight. Its pairing will be changed from flight to flight in a predetermined manner. Thus, a measure of the tracking error will be obtained for each flight. The next set of flights will not be made until the results of the tracking error from the previous flight are known.

If it is not possible to track three balloons simultaneously, the experiment will be modified so as to release two pairs of balloons one hour apart. In this way a 10,000-gram balloon will be released simultaneously with a 350-gram balloon, and an hour later two 350-gram balloons will be released at the same time. Four radio-direction-finding sets will be used to track each pair of balloons, two sets tracking the same balloon. This latter method will increase the size of the experiment. The balloons will be tracked to burst. To keep the time and space variability to a minimum, the balloons will have the same ascent rate. Table 1 shows the type of data that will be obtained from each set of flights.

TABLE I

VECTOR VELOCITIES

FLIGHT NO. / HEIGHT IN FT	1	2	3	......	J	......	N
1000	$\vec{L}_{11}\vec{S}_{11}\vec{T}_{11}\vec{E}_{11}$	$\vec{L}_{12}\vec{S}_{12}\vec{T}_{12}\vec{E}_{12}$			$\vec{L}_{1j}\vec{S}_{1j}\vec{T}_{1j}\vec{E}_{1j}$		$\vec{L}_{1n}\vec{S}_{1n}\vec{T}_{1n}\vec{E}_{1n}$
2000	$\vec{L}_{21}\vec{S}_{21}\vec{T}_{21}\vec{E}_{21}$	$\vec{L}_{22}\vec{S}_{22}\vec{T}_{22}\vec{E}_{22}$			$\vec{L}_{2j}\vec{S}_{2j}\vec{T}_{2j}\vec{E}_{2j}$		$\vec{L}_{2n}\vec{S}_{2n}\vec{T}_{2n}\vec{E}_{2n}$
3000							
4000							
5000							
6000							
7000							
i	$\vec{L}_{i1}\vec{S}_{i1}\vec{T}_{i1}\vec{E}_{i1}$	$\vec{L}_{i2}\vec{S}_{i2}\vec{T}_{i2}\vec{E}_{i2}$			$\vec{L}_{ij}\vec{S}_{ij}\vec{T}_{ij}\vec{E}_{ij}$		$\vec{L}_{in}\vec{S}_{in}\vec{T}_{in}\vec{E}_{in}$
40000							

$\vec{L}_{ij}$ = VECTOR WIND VELOCITY AT LEVEL i OBTAINED BY TRACKING THE JTH 10,000 - GRAM BALLOON

$\vec{S}_{ij}$ AND $\vec{T}_{ij}$ = VECTOR WIND VELOCITIES AT LEVEL i OBTAINED BY TRACKING THE JTH 350 - GRAM BALLOON

$\vec{E}_{ij}$ = VECTOR WIND VELOCITY OBTAINED BY GMD-1 USED AS CONTROL

DATA ANALYSIS. Assuming that the statistics $|\Delta V|_{LS}$ and $|\Delta V|_{ST}$ (shown in Table 2) are random samples drawn from normally distributed populations, the mean value of these statistics for a series of flights will be computed and the difference between these two means will be tested for significance. The 95-percent confidence level will be used for all significant tests.

To be sure that the data are not biased, they will be examined for any departure from normality. There are two ways in which the data could be biased. First, a trend of the vector velocity differences with height for the two 350-gram balloons would be an indication of a within-flight bias. Secondly, different weather conditions could significantly affect the vector velocity differences. This may be called a between-flight bias.

Previous experiments have indicated that the natural variability of the wind does not change significantly with height. The vector velocity differences between balloons of the same type ($|\Delta V|_{ST}$) will give a measure of the natural variability plus the instrumental error. Therefore, any systematic increase in the velocity differences will be due to the tracking system. This error will be calculated by the method outlined in Progress Report No. 138-05, prepared by the New York University under Contract DA 36-039 SC-72, and eliminated from the data. The within-flight bias will be minimised by performing the experiments under approximately the same weather conditions.

TABLE II VECTOR VELOCITY DIFFERENCE

FLIGHT NO. / HEIGHT IN FT.	1	1'	2	2'	......	J	J'	......	N	N'	......	UNPRIMED COLUMNS TOTAL	UNPRIMED COLUMNS MEAN	PRIMED COLUMNS TOTAL	PRIMED COLUMNS MEAN	DIFFERENCE BETWEEN MEANS
1000	$\lvert \vec{L}_{11}\ \vec{S}_{11} \rvert$	$\lvert \vec{S}_{11'}, \vec{T}_{11'} \rvert$				$\lvert \vec{L}_{1j}\ \vec{S}_{1j} \rvert$	$\lvert \vec{S}_{1j}, \vec{T}_{1j'} \rvert$					$\sum_{j=1}^{n} \lvert \vec{L}-\vec{S} \rvert_{1j}$	$\overline{\lvert L-S \rvert}_{1j}$	$\sum_{j=1}^{n} \lvert \vec{S}-\vec{T} \rvert_{1j'}$	$\overline{\lvert S-T \rvert}_{1j}$	$\overline{\lvert \Delta V \rvert}_{LS_{1j}} - \overline{\lvert \Delta V \rvert}_{ST_{1j}}$
2000	$\lvert \vec{L}_{21}\ \vec{S}_{21} \rvert$	$\lvert \vec{S}_{21'}, \vec{T}_{21'} \rvert$				$\lvert \vec{L}_{2j}\ \vec{S}_{2j} \rvert$	$\lvert \vec{S}_{2j}, \vec{T}_{2j'} \rvert$					$\sum_{j=1}^{n} \lvert \vec{L}-\vec{S} \rvert_{2j}$	$\overline{\lvert L-S \rvert}_{2j}$	$\sum_{j=1}^{n} \lvert \vec{S}-\vec{T} \rvert_{2j}$	$\overline{\lvert S-T \rvert}_{2j}$	$\overline{\lvert \Delta V \rvert}_{LS_{2j}} - \overline{\lvert \Delta V \rvert}_{ST_{2j}}$
3000																
i	$\lvert \vec{L}_{i1} - \vec{S}_{i1} \rvert$	$\lvert \vec{S}_{i1'}, \vec{T}_{i1'} \rvert$				$\lvert \vec{L}_{ij} - \vec{S}_{ij} \rvert$	$\lvert \vec{S}_{ij}, \vec{T}_{ij'} \rvert$					$\sum_{j=1}^{n} \lvert \vec{L}-\vec{S} \rvert_{ij}$	$\overline{\lvert L-S \rvert}_{ij}$	$\sum_{j=1}^{n} \lvert \vec{S}-\vec{T} \rvert_{ij'}$	$\overline{\lvert S-T \rvert}_{ij}$	$\overline{\lvert \Delta V \rvert}_{LS_{ij}} - \overline{\lvert \Delta V \rvert}_{ST_{ij}}$
40000																
	$\sum_{i=1}^{n} \lvert \vec{L}-\vec{S} \rvert_{i1}$					$\sum_{i=1}^{n} \lvert \vec{L}-\vec{S} \rvert_{ij}$						$\sum_{i=1}^{n} \sum_{j=1}^{n} \lvert L-S \rvert_{ij}$	$\overline{\Delta V}_{LS_{ij}}$			$\overline{\lvert \Delta V \rvert}_{LS_{ij}} - \overline{\lvert \Delta V \rvert}_{ST_{ij}}$
TOTAL		$\sum_{i=1}^{n} \lvert \vec{S}-\vec{T} \rvert_{i1'}$					$\sum_{i=1}^{n} \lvert \vec{S}-\vec{T} \rvert_{ij'}$							$\sum_{i=1}^{n} \sum_{j=1}^{n} \lvert \vec{S}-\vec{T} \rvert_{ij'}$	$\overline{\Delta V}_{ST_{ij'}}$	

$\vec{L}, \vec{S}, \vec{T}$ ARE SAME AS IN TABLE I

$$\lvert \Delta V_{LS_{ij}} \rvert = \lvert \vec{L}_{iij} - \vec{S}_{ij} \rvert = \lvert \vec{L} - \vec{S} \rvert_{ij}$$

$$\lvert \Delta V_{ST_{ij}} \rvert = \lvert \vec{S}_{ij} - \vec{T}_{ij} \rvert = \lvert \vec{S} - \vec{T} \rvert_{ij}$$

SIZE OF EXPERIMENT. The sample size that will be necessary to determine whether of not the mean of the difference between velocity difference ($\overline{|\Delta V_{LS} - \Delta V_{ST}|}$) will not differ from the true value of the mean by more than one-half mile per hour can be calculated. Assume the following condition:

1. The standard deviation of the sample is equal to the population standard deviation. (From prior knowledge of the probable error of the tracking system and the natural variability, the standard deviation is estimated to be 4.5 mph.)

2. The velocity differences $|\Delta V_{LS} - \Delta V_{ST}|$ are normally distributed.

If we let x equal the true magnitude of the vector velocity differences (the meand that would be obtained from an infinite sample), then the mean of a finite sample $\overline{|\Delta V_{LS} - \Delta V_{ST}|}$ will be normally distributed with a mean x and a standard deviation $4.5/\sqrt{n}$. To be reasonably certain that $\overline{|\Delta V_{LS} - \Delta V_{ST}|}$ will be within one-half mph of x, 95 percent of the time, let one-half mph = two standard deviations or $9/\sqrt{n}$ and solve for n. In this case n = 324. Therefore, at least ten flights of three balloons released simultaneously (one 10,000-gram balloon with two 35-gram balloons) will be needed to obtain an accuracy of one-half mph in the mean.

MAN-HOUR REQUIREMENTS. The experiment will require personnel to maintain and operate four radio-direction-finding sets (GMD-1), to inflate the balloons, and to analyze the data. Tables 3 and 4 indicate the approximate number of man-hours required to complete the experiment. Approximately 45 man-hours will be needed for conducting the flights and 140 man-hours for analyzing the data and preparing the report.

CONCLUSIONS. Three conclusions can be drawn from this experiment:

1. $\overline{|\Delta V_{LS}|} - \overline{|\Delta V_{ST}|}$ does not differ significantly from zero. This indicates that for the degree of accuracy desired in wind velocity measurements, both the 350-gram and the 10,000-gram balloons follow the wind equally well.

2. $\overline{|\Delta V_{LS}|}$ is significantly greater than $\overline{|\Delta V_{ST}|}$. This seems to indicate that the 10,000-gram balloon is either more sensitive to instantaneous changes in the wind and, therefore, gives a better measure of the true mean wind, or, due to its size, measures wind to a different scale.

3. $\overline{|\Delta V_{ST}|}$ is significantly greater than $\overline{|\Delta V_{LS}|}$. This indicates that the 350-gram balloon is more sensitive to instantaneous wind changes and thus gives a measure of the wind velocity to a smaller scale.

TABLE 3

MAN HOUR REQUIREMENTS FOR EXPERIMENT

OPERATION	PERSONNEL	TIME FOR EACH OPERATION	MAN HOURS FOR 10 SETS OF FLIGHTS
INFLATION OF 10,000-GRAM BALLOON AND TWO 350-GRAM BALLOONS	1	1 HOUR	10 HOURS
RADIOSONDE PREPARATION	3	$\frac{1}{3}$ HOUR	10 HOURS
BALLOON RELEASES	5	$\frac{1}{5}$ HOUR	10 HOURS
TRACKING TO BURST	1	$1\frac{1}{2}$	15 HOURS
TOTAL	5		45 HOURS

TABLE 4

DATA REDUCTION AND ANALYSIS

STEPS IN ANALYSIS	PERSONNEL	TIME REQUIRED FOR EACH STEP	MAN HOURS FOR 10 SETS OF FLIGHTS
EVALUATION OF RADIOSONDE	1	1 HOUR	30 HOURS
WIND VELOCITY COMPUTATIONS	1	1 HOUR	30 HOURS
STATISTICAL ANALYSIS	1	8 HOURS	80 HOURS
TOTAL			140 HOURS

TABLE 4

DATA REDUCTION AND ANALYSIS

STEPS IN ANALYSIS	PERSONNEL	TIME REQUIRED FOR EACH STEP	MAN HOURS FOR 10 SETS OF FLIGHTS
EVALUATION OF RADIOSONDE	1	1 HOUR	30 HOURS
WIND VELOCITY COMPUTATIONS	1	1 HOUR	30 HOURS
STATISTICAL ANALYSIS	1	8 HOURS	80 HOURS
TOTAL			140 HOURS

EVALUATION OF INFECTIVE VIRUS PREPARATIONS AS TO POTENCY

F. M. Wadley
Program Research Branch, Assessment Division
Fort Detrick, Maryland

THE PROBLEM. Owing to the nature of pathogenic viruses, the techniques used for population estimation with bacteria are impossible. Virus particles can be seen only with difficulty by the use of electron microscopes, and direct counts are out of the question. Such methods as tissue culture or complement-fixation from dillutions have some use, but are expensive and slow. Reaction of hosts to injections from successive dilutions is the method which is usually employed. Units of response must be defined in an objective way, as will be discussed below.

ANIMAL RESPONSES. Responses are divided logically into two kinds, graded responses and all-or-none responses. With graded response, a reading is taken from each individual subject. The time from injection to onset of symptoms is a typical graded response. With all-or-none responses, the only record made for a host is that it did or did not respond; it was diseased or not diseased, died or survived. In either case, responses from successive concentrations are analyzed by regression methods for estimation. Approximate parallelism in regressions is required for comparison of two or more materials in the most effective way.

A successful graded response is more precise than an all-or-none response. Its connection with basic regression procedure is simpler. In virus work in our laboratories, graded responses have not been especially successful. Such responses as time-to-death, weight loss after dosing, and time-to-onset of fever have been used. The relation of these measures to concentration has not proved as strong as is needed in a good graded response.

The all-or-none tests principally have been employed in our laboratories and have proved quite workable. They are analyzed fitting dosage-mortality or dosage-effect curves with appropriate modifications of regression prodedure. The common log-probit treatment of dosage-mortality curves has been used. Several other treatments might be employed with similar results. Mice have been used for the injection work.

PROCEDURE. Test animals are injected in a standard manner, using successive dilutions of the virus. Selection of a series of concentrations or dilutions is dependent upon expected slope and position of the curve. With some knowledge of the concentration which will give partial response, and a fairly steep dosage-mortality slope, half-log dilutions may be employed. With less knowledge of favorable concentration, or with slight dosage-mortality slope, 1-log concentration intervals are used and a wider range is explored. It is desirable to have partial mortality at 2 or more concentrations; zero and 100% mortalities have only limited usefulness. As few as 3 or 4 concentrations may give good results if some preliminary knowledge is available. No phase of experimental design is more taxing to the biometrician than selection of concentrations.

Injected animals are held for the required period, to observe and record response or non-response. The percentages of response are transformed to probits, while logarithms of concentration or dilution are taken. Dilutions are usually stated as powers of 10, so that the stated exponent is the log sought.

These transormed data are used in standard probit analysis (Finney, 1952, "Probit Analysis" describes this adequately). The concentration bringing about a 50% response (ED_{50} or LD_{50}), the probit slope, variances for these quantities, and other estimates may be calculated.

From the ED_{50} we may estimate some quantity such as "mouse units" (the amount of material per mouse required to bring about a response in half the subjects). The number of such units per unit volume of virus stock may be estimated and used in later steps. To this point, no special complications are found. We have defined a unit of concentration, although we cannot state it in virus population terms. In probit analysis, it is supposed that concentration is known rather exactly, though response varies. With chemical toxicants there is no doubt of precise knowledge of concentration; with biological materials the assumption may not hold.

However, in the stage of dosage-mortality calculations with dilutions of virus, the precision of dilution is not a pressing problem. The unknown concentration is used in dilutions carefully carried out. While it has been shown in special studies that dilution and other phases of technique are subject to errors, they do not seem likely to be large compared to variation in the response of limited numbers of animals. It is later, in using the estimates of concentration secured from dosage-mortality study in further work, that error in concentration becomes important.

PRECISION AS COMPARED WITH BACTERIAL PLATE COUNTS. Precision of estimates is measured by the variance of the log ED_{50} from the probit analysis. This measure is affected by the total number of subjects used, their probit weighting (greatest near 50%), the probit slope and the nearness of log ED_{50} to mean log concentration (see Finney l.c.). The second and fourth factors are modified by careful placing of concentrations. As to probit slope, it is greater with more uniform results and lower with variable results. In tests of this sort, a slope of 2 or more is considered good, and a slope of 1 is regarded as poor. Given fairly well-managed experimental conditions, the total number used is the principal factor in variance of log ED_{50}.

Using this variance, confidence limits are worked out for log ED_{50}, and limits are transformed to antilog or concentration terms. On the log scale the limits are additive; on the concentration scale they become multiplicative. Some recent tests may be cited.

In one virus study (Mr. W. C. Patrick of our laboratories), extensive injection tests were conducted, with the object of improving technique and controlling quality. A sample of 15 recent tests was studied. In each, 40 mice were used; 10 mice at each of 4 dilutions. Dilutions were grouped around the expected ED_{50} in half-log intervals. Slopes were fairly good - over 2 on the average. In two of the 15 tests, only one partial response occurred; the other responses were 100% or zero. These results could be used in interpolation, but were of little use in probit analysis. In three

other tests, variance was high, because of wide scatter leading to a significant chi-square, or because of wide separation of log ED_{50} from mean log concentration. In these three tests, confidence limits averaged about plus or minus 1 log (ten-fold on the concentration scale). In the other 10 cases, operation was smooth, and confidence limits on the log scale were from 0.16 to 0.52, averaging 0.26. This is equivalent to "times or divided by" 1.82, or about 80%.

In more extensive evaluations in a recent experiment (Mr. George Harris), 160 mice were used, 40 at each of 4 half-log dilutions. In five such tests, one showed a significant chi-square with confidence limits of plus or minus 0.64 on the log scale (about four-fold on the concentration scale). With the other four, confidence limits varied from 0.11 to 0.17, averaging 0.14. This value is equivalent to "times or divided by" 1.4 or about 40%.

These typical cases give an idea of precision to be expected under fairly good conditions. With moderate numbers of mice, a minor fraction of tests may miss the mark; the more successful tests may estimate the desired point within less than a two-fold range. Quadrupling the number of animals will approximately halve the limits.

Bacterial plating to estimate populations has a much lower variance. Plating technique aims at getting 100 to 300 colonies per plate, for the sake of precision and adaptability to counting. Several plates are used for one estimate. A typical series of actual plate counts, all from the same material, is 198, 151, 180, 184 (Mrs. Claire Cox's data). Using these and several other similar sets, with 22 degrees of freedom within sets, a variance of 269 is calculated. Transformation is not needed under the conditions dealt with here. The 95% confidence limits for the average of 3 plates are plus or minus 19, which is about 11% of the mean (174 for all sets). Plating variance at its lowest will approximate the mean (Poisson condition), but it is usually somewhat higher, as in the case cited. Comparison of bacterial and viral precision is made for situations in which considerable attention is given to technique, with organisms which respond well. It should, thus, adequately represent relative precision.

<u>PROCEDURE IN STUDYING AEROSOL RESULTS</u>. Samples of aerosol are taken at several time intervals and collected in liquid. Injections of this liquid into mice at several suitable dilutions are carried out, and a dosage-effect curve is derived for each. In each case, the ED_{50} may be used to estimate log of "mouse units" per liter. These estimates at each period are plotted against age of aerosol in minutes. The relation is presumed to be linear, and does in fact seem close to linearity. A line is fitted; the regression coefficient gives an estimate of decline of concentration with time, on the log scale. By taking the antilog (2 minus the coefficient) a percent remaining is secured. Subtracting it from 100 the "decay rate" in percent per minute is estimated.

By extrapolation backward, an estimate of log concentration at zero time is secured. The first sampling is made only a few minutes after the start, and the value for the estimated intercept is usually close to that from the first sample. By means of estimates made from the stock material

before aerosol formation, and from dilution factors, an estimate of expected log concentration (with perfect success in aerosolization) may be made. The difference between actual and expected is the log loss. By means of the sort of antilog procedure used in studies of decay rate, an "initial percentage recovery" or a "source strength" is estimated.

Initial percentage recovery and decay rate are estimates much used in bacterial and viral aerosol studies. It is obvious that initial percentage recovery of virus has a high sampling error. The pre-aerosol concentration is estimated, subject to variation in animal responses. Aerosolization undoubtedly has some real variation. Aerosol concentration therefore must again be estimated from animals. These three tandem sources of variation make percentage recovery estimates quite variable. Decay rate estimation involves only the third source named, and is more stable. Final sampling error of the two estimates is calculated from variation in repeated trials.

Respiratory effect of aerosols is tested by exposing animals for varying periods at varying ages of the aerosol. Log dose in "mouse units" unhaled is estimated from concentration of the aerosol at the given time, exposure time and breathing capacity of the animal. Dosage-effect curves can be fitted, and estimates can be made of respiratory ED_{50} in "mouse units", of slope and other factors.

This last stage - fitting dosage-effect curves using concentration estimates derived from other similar curves - very definitely involves variation is estimation of dose. This variance can be estimated from the dosage-effect curves of injected aerosols. The ordinary probit fitting procedure must be modified, using this estimated variance. A study made in our laboratories (SB Report 1773) is used in the modification.

The variable estimate of dose does not lead to bias in the estimates of ED_{50}, et cetera, but does increase their variance. The situation can be dealt with (in the probit solution) by appropriate decrease in the weights. The reciprocal of one plus the product of three factors, the variance of dose, the square of the estimated regression coefficient, and the standard weight, is used as a multiplying factor to reduce the weight. Nomograms have been developed which can be used to shorten the calculations. The variance estimate for ED_{50} may easily be doubled by considering dose variance.

Errors of initial recovery and decay rates are calculated on the log scale from determinations in replicated trials. Confidence limits are translated to the concentration scale. With the variation developed in the involved procedure necessary, individual aerosol dosage-effect trials with small numbers of animals have often failed to yield data adequate for probit analysis. When values for several trials were put together, fairly good probit analysis was possible, and yielded ED_{50} values and probit slopes with their confidence limits. It would be an improvement if larger numbers in individual trials could be used, and if these confidence limits could be calculated from variation in repeated determinations.

This discussion has presented little of experimental design in the narrow sense; only the simpler designs have proved usable. Completely

random designs, or where two or more methods are compared - randomized block analogues, are the usable plans. More complex designs might come later. The broader and more fundamental aspects of design, definition of objectives, planning of valid comparisons, and utilization of previous information in planning have been of great importance. Selection of dilutions and of sampling intervals for aerosols are especially important.

EXPERIMENTAL DESIGN FOR FIELD STUDIES IN LEADERSHIP*

Carl J. Lange and Francis H. Palmer
Human Resources Research Office

The problem of studying leadership can be viewed as part of the general problem of studying social interaction with special emphasis on social influence processes. A recent trend in theories of personality emphasizes the inter-personal nature of human behavior; it has had a parallel in studies specifically related to leadership. The trend has been away from a search for traits differentiating leaders from nonleaders, and toward analysis of the interaction among leader, situational, and follower characteristics.

With the recognition of the importance of studying leadership from an interactional point of view has come a new emphasis on the use of field studies. There are distinct advantages in studying problems of leadership in real rather than simulated situations. The multitude of complex variables, and, in particular, motivational variables, that form the context for human behavior in groups cannot easily be duplicated in contrived situations. The ecological validity obtained with samples of real groups functioning in real settings enables generalization with considerably greater confidence than is possible when simulated groups are used.

Of the two major types of field study, exploratory and hypothesis-testing, the former is used when little information exists about the nature of the group or the activities in which they are involved, or when the information is such that it does not yield logical and clearly-defined hypotheses. A correlational design is often appropriate for such an initial investigation to determine which variables are of primary importance and to provide empirical bases for the formulation and refinement of hypotheses. In an area as complex as leadership practical administrative requirements necessary to achieve technical requirements of complex experimental designs frequently make the adoption of correlation design better strategy in early stages of research.

Two exploratory field studies on leadership utilizing correlational designs which have been conducted at HumRRO's Fort Ord research unit will be discussed here. The first, Palmer and Myers' study of human factors contributing to the productivity of antiaircraft batteries,**focussed on the group and its performance on critical activities, relegating leader

---

* The research reported here was conducted by the authors while employed by the George Washington University, Human Resources Research Office, operating under contract with the Department of the Army. Opinions and conclusions are those of the authors and should not be construed as representing those of the Department of the Army.

** Hereafter referred to as the AAA study.

characteristics and behavior to a role no more important than that of various other factors which might reasonably be related to productivity. In discussing this study, emphasis will be placed on problems pertaining to criterion measures of group productivity and interpretation of results.

The approach of the AAA study, with its interest in identifying variables which were closely related to productivity rather than in group leadership *per se*, permitted comparison of the leader's influence upon productivity with that of other variables. In correlational productivity studies, the problem of reliability is especially critical and concerns both measures of performance and measures of group characteristics. The nature of measures associated with behavioral phenomena in the field makes reliability a significantly greater methodological problem than is often true in laboratory situations where the experimenters' experience and care are major considerations. Considerations of validity differ with respect to the two kinds of measures: for validity of criteria, experts in the performance area involved must necessarily participate in the selection; for validity of group measures, the investigator must rely on his own knowledge and professional background. When both criterion and group characteristic measures have been fixed, and the data collected, the most severe limitation of correlational design becomes evident: the interpretation of the resulting matrices of correlation coefficients. If the study is large several thousand coefficients may be available, and since many of the variables are likely to be related, it is difficult to determine the precise number of relationships expected by chance alone for a specified level of confidence. Thus, the selection of relationships as the basis for hypotheses warranting subsequent testing requires judgments which should be made with extreme care. How these problems were dealt with in the AAA study will now be described.

Forty antiaircraft batteries in a single defense were studied. As far as could be determined the assignment of personnel to these units had been random. The equipment they used was for practical purposes identical, and, of course, each battery had an identical mission. A poll of senior officers of seven AAA defenses was taken to assist in identifying criteria. There was high agreement on three primary activities upon which achievement of a unit's mission depended. Developing measures for these activities was, of course, complex, and the results showed that the measures finally evolved were of varied accuracy. One (the range of radar pickup) proved reliable at the level of .87; the reliability of the scores for the second measure, readiness to engage target, was so low that the measure could not be used in the analysis; in the case of the third (equipment maintenance scores) the data did not lend themselves to any satisfactory treatment to determine reliability.

The measures of battery characteristics included intelligence, education, personality, leadership, group structure, life history information, and sociometric data. The resulting scores were grouped in several ways. Means and variances for these measures were determined for the battery as a whole, for each sub-section, by rank and position, as well as for certain smaller categories. Leaders' characteristics for the battery commander, the first sergeant, and the several section leaders were treated separately.

Over 100 measures, grouped in these categories, were related to Range

of Radar Pick-up and maintenance scores, and to three criteria of secondary importance to the study. Several matrices, representing several thousand coefficients, were available for interpretation. Since a number of relationships would be expected to reach statistical significance as a function of chance alone, certain criteria were developed for selecting those relationships which would be used in subsequent research. First, it was decided to consider only coefficients significant at the .05 level or better; second, the relationship must make some kind of psychological sense; and third, it had to hang together with other variables known to be interrelated. In this manner, clusters of relationships were given more weight than isolated relationships which did not seem to go with other results found in the study. With the application of these criteria to the data, many meaningful hypothese were derived from the study. To be sure, before anyone made too much of the correlational data, the particular relationship concerned would have to be validated. However, as a source of hypotheses about the relationships between group characteristics and productivity, including the influence of various levels of leadership, the investigators considered the study extremely valuable.

The second field study of leadership to be discussed here was concerned with identifying actual on-the-job behaviors which differentiate between effective and ineffective leaders. As part of a long-range program of research whose ultimate goal is to provide validated leadership doctrine for use in training Army officers, our early efforts have been directed toward the problem of identifying leader behaviors which correlated with evaluations of leader effectiveness. The discussion of this second study will emphasize the steps taken to obtain objective, bias-free estimates of the occurrence of various types of leader behavior within the framework of a field study.

Perhaps the major methodological problem involved in field studies of this type relates to the provision for control in the observation of behavior and in the recording and analysis of data. There are numerous obstacles involved in providing the necessary control. One such obstacle, made famous by the Hawthorne studies, is that of obtaining data without influencing the groups studied. The presence of observers scrutinizing the activities of the group may have unknown effects on the performance that could make findings fallacious when applied to unobserved groups. Obtaining retrospective reports from members of the group is one practical method of overcoming this obstacle.

In our study, we were interested in obtaining accurate descriptions of overt behavior on the part of the leader in certain specified types of situations. We were especially interested in verbal communication behavior. This interest is in contrast to research which is primarily concerned with communication structure or power structure, that is, research that asks questions such as "Who interacts or communicates with whom and how frequently?" or "Who influences whom?" Our question was, "What are the differences in verbal communication content between those who are accepted and those not accepted by followers?" Accordingly we needed accurate and detailed accounts of the leader's verbal communication behavior and other overt behavior in certain prescribed situations.

The question arises, "Are retrospective reports of group members

which describe leader behavior detailed and accurate enough to be useful?" To answer this question, we did a small trial study in which we obtained retrospective reports of previously observed behavior, and in the same interview, sensitized the reporter to the kinds of observations we desired. Then, in subsequent interviews spaced one week apart, we obtained additional descriptions of newly observed behavior from the sensitized observer. A comparison of the unsensitized and sensitized reports revealed the information obtained in the initial interview was as detailed, specific and abundant as that obtained in later interviews. Thus, we were able to obtain observations of leader behavior as it occurred in natural, representative situations.

Now, I would like to discuss procedures we used in getting measures of leader behavior with special emphasis on steps taken to eliminate sources of systematic error that can easily creep into studies of this type.

First, let me elaborate on the method used for obtaining the basic data, i.e. the descriptions of the leaders' behavior in specified situations. An interview technique was used. The interviews with the leaders' subordinates features a standard set of questions aimed at getting an exhaustive description of the leader's behavior based on retrospective eye witness accounts. We asked the respondants to report actual incidents, and a heavy emphasis was placed on getting behavioral, rather than inferential reporting. The situations included:

1. Job assigning or planning
2. Job in process and being done poorly
3. Job in process and being done well
4. Job completed and done poorly
5. Job completed and done well
6. New men entering group
7. Promotions or changes in assignment
8. Group members making complaints or suggestions
9. Unexpected events occurring.

In the interview, no evaluative comments were asked for; interviewers were carefully trained to encourage specific and detailed reporting but to avoid reacting differentially to particular types of information.

Several advantages of this approach to obraining behavior descriptions may be noted. First, aside from the types of situations used, no restriction is placed on the content of behavior reported as contrasted to the commonly used questionnaire approach, wherein the particular behaviors about which information is obtained are predetermined and limited

by the investigator. A second advantage accrues from getting descriptions of specific occurrences of behavior. With the use of questionnaires, the task of the observer is considerably more complex in that he must interpret the statement on the questionnaire and integrate his past observations to arrive at a summary response. He has the burden of selecting, weighting, and summarizing observations. The rules used by the various observers are not explicit, and there can be little certainty about what actual behaviors determined the response.

Additional precautions were taken in collecting data to avoid systematic error. In our design, we needed ratings of the general effectiveness of the leaders studied. The interviewers who collected descriptions of the behavior observations knew nothing of the evaluative ratings of the leader. Also, the followers of each leader were distributed among the four interviewers to prevent impressions gained in one interview from influencing interviewer behavior in subsequent interviews, and also to avoid systematic effect of particular interviewers.

Having obtained descriptions of observed leader behavior reasonably free of distortions and restrictions, we had the problem of translating these qualitative descriptions into quantitative scores that would be useful. We developed a set of categories which included both situational or contextual information and behavior information. The situational categories included such information as location, type of task or activity, stage of task, (i.e. beginning, in process, or completed,) importance and rountineness of task, person or persons with whom the leader is interacting, and other persons observing the interaction. (e.g. the presence of "brass" in a particular situation could be recorded.) The behavior categories fell in five main areas: (1) Defining (2) Motivating Performance (3) Handling Disrupting Influences (4) Getting Information (5) Uses and Support of Subordinate leaders.

The completed set of scoring categories included roughly 140 dimensions of behavior or situational context, each dimension having from 2 to 10 quantitative or qualitative alternative scores. The entire list of categories was applied to each scorable unit of interview data, a scorable unit being, in general, a single scene or incident of the leader interacting with group members.

Each of the variables was defined in terms of overt, observable characteristics. Classifications of behaviors were not made on the basis of inference about the leader's intent or about the probably effect on group members. A scoring manual was prepared which objectively defined each scoring category, defined a symbology for scoring, and laid down a set of general scoring instructions and limitations, aimed primarily at preventing subjective inferences being made by scorers. Six trained scorers categorized all the interview data, the interview data for a given leader being distributed pretty evenly among all scorers.

Final scores used in the analysis were obtained by weighting raw frequencies of a given behavior by the total amount of data provided for a leader. Many of the basic behavior category arrays were arithmetically combined to create more general behavior variables.

This approach to the analysis of the qualitative data provided scores that were derived by applying explicitly defined operations. Scorers' judgments were limited to single items of information. No subjective summarizations or integration of the data were made. And, thus, the effects of various types of scorer bias were minimized.

In describing the method we used to obtain leader behavior scores, I have emphasized the steps we took to insure that our measures would be free from bias of various types and to make explicit each operation used for obtaining the scores. Having taken these steps, we have measures of general behavior variables which are explicitly tied to day to day observations of specific behaviors.

These measures were related to superiors and subordinates' evaluations. Since the score distribution of the criterion measures were roughly bell-shaped, we decided to employ Pearson r as an estimate whenever the behavior variable distributions were bell-shaped and continuous. For those behavior variables which did not meet this requirement, Chi square was used. The resulting analysis provided information about the relationship between leader behavior variables and effectiveness ratings of the leader by subordinates and superiors.

In summary, two field studies using correlational design have been discussed with special emphasis on methodological problems commonly faced.

THE DESIGN OF CONTROLLED FIELD EXPERIMENTS

Floyd I. Hill
Technical Operations, Incorporated

I. INTRODUCTION. The Combat Development Experimentation Center at Fort Ord, California, has the mission of examining Army organizations, procedures and doctrines experimentally. To accomplish this, it has available 3,000 troops and approximately 250 square miles at Hunter Liggett and Camp Roberts Military Reservations. Supporting this operation is a Research Office consisting of approximately twenty professional and twenty-five support personnel whose mission is to advise the Commanding General of CDEC on the design, conduct, and analysis of these controlled field experiments. CDEC is relatively new, having been established on 1 November 1956. However, its experimental activity has been at a high level since the beginning of the first experiments in March of this year.

The subject matter of this paper can be divided into three general headings. The first is on the need and nature of controlled field experiments. The second is a discussion of how we arrive at our experimental designs within the specific set of limitations imposed by our operation. The third is a discussion of some sample designs.

II. THE NEED FOR AND NATURE OF CONTROLLED FIELD EXPERIMENTS. Under the impact of a rapidly changing technology, it bacame apparent some years ago, that the Army should establish an organization devoted to the development of new military organizations, tactics, and doctrines capable of utilizing the results of our new technological efforts. The mission of Combat Developments is a complex one. It must devise new organizational structures, and, if necessary, new operating tactics and doctrines. Some of this can be done by analysis, particularly if no major organizational change is made or no major doctrinal change is made. That is, perturbations of earlier solutions are likely to be successful. However, major departures from tried and true solutions in both these fields are associated with the uncertainty of extrapolation. New organizations interact with new tactical doctrines. Thus, a new table of organization and equipment is very likely to require new tactics for effective operation.

Indeed, the approach toward organizational and procedural experimentation has been principally "proof of the pudding" or field tests. The question that has arisen as a result of single field tests with radically changed organizations and procedures is, "Do these single tests prove anything?" Another is, "To what extent were they influenced by a small group of individuals or by changes in the individuals' activities?"

The student of organizations and tactics is attracted, at first, to a concept of studying small unit sub-organizations in a highly controlled environment. With these small units, he would like to develop some sub-tactics, to coin a word, that will go with these sub-units, and then, by some combinatorial technique, predict the effectiveness of a larger organization. Such an approach must be made with caution since it tends to violate the basic premise of the organizational arguement. That is, that there is a uniqueness associated with combinations of men

and weapons rather than there being a simple combination of sub-components derivable from sub-component performance that will give an effectiveness measure of the higher component. If the combinations are quite complex, as we suspect them to be, additional experimentation would have to be conducted to determine the nature of these combinations, for surely it is not understood at present.

For this reason, the analysis in Combat Developments has largely followed the war game technique. That is, the organization is looked on as a whole in the environment in which it was expected to perform. These war games are characterized by two sidedness, i.e., the interaction between friend and foe, and by mutual support, i.e., the interaction between supporting and associated elements. Thus in military organizations, we are interested in the squad as a part of the platoon, and the platoon as a part of the company, and so forth.

Three types of war games are used in the evaluations. The first might be characterized as the "paper" war game, where antagonists using separate rooms, battle it out on a map. The second is the "machine" war game, presently in a state of development for land combat, which has the advantage of speed necessary to arrive at Monte Carol type solutions to the war game. The third is the "field" war game, where all the elements of combat are present, consistant with safety of the players.

Each of these techniques supports the other. Both the machine and the paper war game demand a type of data in short supply. These data are those associated with the response time of the combatants to enemy or command action, the space time, or movements of the forces under conditions of fire, and target characteristics of the forces, i.e., their density of disposition, cover, etc. The controlled field experiment is designed to produce just such data.

The limitations of the controlled field experiment are that it can use only equipment which exists or can be simulated simply, and it is expensive in men and time. Thus, while we are limited with the paper war game to tactics of division and corps sized forces, and with the machine war game to a distribution of solutions of a single problem, they are strong where the field experiment is weak and vice versa.

With the data of the controlled field experiment, there can be the expectation of effective machine analysis, and with more effective machine analysis, paper war games promise to produce more reliable results. In turn more meaningful field experiments can be designed based on the results of more meaningful war games.

III. THE EXPERIMENTAL DESIGN PROBLEM.

As with any experimental design problem, that of CDEC has those characteristics of uniqueness and generality. Perhaps the problem most general to all experiments is that of the measure of effectiveness. In testing an organization or a tactic, we ask the question, "What is the criterion of goodness?" In some experimental situations, the criterion appears quite simple. However, in the great majority, simple criteria are achieved at the expense of the questionable validity of these criteria. If we examine the organizational or tactical problem, it is

apparent that the measure of effectiveness must be associated with the timeliness of the accomplishment of the mission assigned, the cost in accomplishing this mission, and the damage inflicted upon the emeny. Simply, this may be stated that the criterion is enemy casualties, friendly casualties, and time of mission accomplishment.

This we may call a multiple factor criterion or, more simply, a three headed monster. A large portion of our effort has been devoted toward reducing this to a single measure of effectiveness. To date, we cannot claim success in the reduction to a single number. We have considered several approaches. First, we have attempted multiple regression analysis, using linear combinations of several characteristics of the times and the number of combatants, and casualties. Our principal difficulties in this area have been in the determination of mission time and the subordinate times which make up the accomplishement of a mission. Currently, we are dealing with sub-sets. We have classified them as approach, development, fire, fight, and assault. A summary of our progress to date in the criterion problem is given in a paper recently given to the Western Section of the Operations Research Society of America in San Francisco, California.*

The criterion of effectiveness is closely associated to the type of experiment we conduct at CDEC. Our effectiveness in estimating the casualties on the basis of weapons effects appears to be better than many other aspects of our experimental control. However, given that we have tools for the allocation of casualties, the effectiveness of our technique depends on the type of experiment that we must run. First of all, this experiment must study a characteristic associated with response time, space time, and target characteristics. We do not attempt to design maximum seeking experiments. Rather, we approach the problem of discrimination among alternatives. We have no immediate hope of experimentally establishing alternatives. Rather, the military presents organizational structures or tactics which they believe to be competitors for discrimination as to which is the best. These candidates are derived under several limitations. First of all, with 3,000 troops and, say, four candidates, there is little likelihood of testing an organization greater than that of a company in size, at the present time. In modern warfare the area that we deal with is scarcely large enough to accommodate actions of a battalion sized force, and with many more troops, it would be difficult to study any organization larger than a battalion. The controlling numbers are surprisingly simple. If we take a company of 250 men and consider four company organizations as candidates made up of different groups of people, 1,000 men are used just in the candidate organizations. The Agressor should be roughly the same size, and umpiring of this organization will require of the order of 300 people. If we consider supporting forces both in terms of military support and those that adminster to the testing organizations, we find our 3,000 men used up. Thus all experiments to date have been with no larger than with conventional size forces.

There are other limitations concerning the stability of the group we deal with. A military force on any one station today is a constantly

---

* Measures of Effectiveness in Controlled Field Experiments -- Presented at Western Section of ORSA, 27 September 1957 -- Floyd I Hill and Walter E. Pearson.

changing group of people. If you desire any stability of command or personnel, the experiment should not last longer than about three months. Even in a three-month long experiment, 20 - 30 per cent turn over of umpires, Aggressors, and candidate organizations can be expected.

The experimental designer is faced with several nuisance factors, in addition to the foregoing limitations. These are principally associated with the fact that in dealing with organizations or tactics, he is dealing with a group of human beings who are inherently different and at different levels of learning. The learning is particularly difficult since these same human beings learn a piece of terrain as they pass over it. The learning of terrain is so important that we believe it almost mandatory that record experimentation be conducted over terrain which has not previously been passed over. By terrain we mean not only the land situations but also the disposition of the enemy on the land. This difficulty is not encountered, of course, on the high speed computer. In addition to these primary nuisance factors, there are secondary ones which influence all land warfare operations, such as weather, morale, and so forth. These cannot be called minor under many conditions. For instance, the Hunter Liggett Military Reservation becomes almost impassable for wheeled or tracked vehicles off the raods once the rains start early in December. Morning fog is frequent, and the temperature range during the summer months between night and late afternoon is of the order of 70°F. Morale can become a major problem if an experiment causes the suspension of the Christmas holidays, and can be overriding in the consideration of the length of the experiment.

One of the most important steps in our experimental procedure is the development of adequate candidates to fulfill the objectives. Almost any experimental requirement presented has a multiplicity of objectives. It is the problem of the soldier-scientist team at CDEC to group these into groups which can be considered to achieve the objectives. For example, if an objective is to determine the control characteristics of a company sized organization, then the candidates might be companies with different numbers of platoons and different headquarters structures. Then we must ask ourselves under what conditions is it likely that control might be stressed. These conditions would include meeting engagements, an attack against a defended position, a defense of a position against an attack, and delaying actions. These would constitute a family of situations each candidate must be considered in. In addition, we must look at the case where interactions are most likely to occur. A major area in considering the control of a company organization is the span of operations. Thus, in a controlled experiment, we are likely to consider the six foregoing types of engagements for each span of operations with each candidate. Thus, the outline for a scenario is prepared and a description of the events occurring in each cell of our experimental plan is developed.

The selection of a candidate is a problem of considerable difficulty. In an experimental procedure demanding discrimination, the scientist asks that the candidates considered be sufficiently different that, if discrimination does not occur, i.e., that they appear similar, important knowledge is gained. Such an approach places a great responsibility not only on the experimental designer, but on the man presenting the candidates for discrimination. We may well ask, "What is a significant difference?"

From our point of view at CDEC, small changes are not of great interest. The type of Army problem we are looking at is large and cannot be solved by small changes in effectiveness. We then may well ask, "What is a large change?" To begin with we start not with the random differences to be anticipated (since our experience does not make us able to do this) but with the sample size itself. If, we might ask, an organization or procedure operating under highly controlled conditions cannot produce a detectable difference in the time of mission accomplishment, enemy casualties, or friendly casualties, in say, four replications, then is a significant difference likely? If four is not enough, then is five, and so on. Thus, one of the major steps in the design of our experiments at CDEC is the development of candidates believed to be, in truth, different. We might ask at this point, as Pilate did, "What is truth?" Perhaps truth is a 20 - 30 per cent difference in combat effectiveness. Surely it isn't less. Thus, we start examining our sample size, that is the number of replications, not on the basis of the variability of our outcome, but rather with the demand that our candidates be sufficiently different that the variability of our outcome be small in comparison. If the variability of our outcome is large, thus demanding a very large number of replications, a military requirement of a reasonable degree of certainty for improvement is not met. With this requirement set upon the number of replications, we can expect that the number of variables we are to study is more likely to control the number of replications in our experiment than the random variation in the outcome.

If we have a set of candidates, these represent one variable. Each of these candidates then must be tested by a leader group (a leader and a group of individuals) operating over a certain piece of terrain at a certain level of learning. This group will be opposed by a given group of Aggressors, who also will be under a given leader at a given level of training. Clearly, the influence of this leader group on the slection of a candidate could be great. Further, it is exceedingly unlikely that this leader group could test one candidate and then be at the same level of training when it tested the next candidate. Also, it is desirable that there be at least as many leader groups as candidates, and further, that each candidate be tested by one of the leader groups at each of the levels of training.

We might suggest that we train a leader group to the point where testing with, say, one organization or tactic would not increase the absolute level of training of this leader group in testing another organiaztion or tactic. Practically, this is very difficult to achieve. Proving ground experience in weapons testing has shown that this level of training is not achieved with men with years of experience in testing a single weapon. In operating organizations where stability for a three-month period is the best that can be expected, no such level of training can be anticipated anyway. One other learning factor gives us trouble. Experience has shown that a leader group tends to learn a particular piece of terrain and how to operate on it very rapidly. Once the group has learned a piece of terrain, its activities are adjusted sufficiently as to not be representative of activities of a group operating on an unfamiliar piece of terrain. Since we expect most military operations of significance to be those conducted on terrain not previously

traversed by the combatants, our experiments are concerned with the performance of individuals in relatively unfamiliar terrain situation. Thus as many terrains as there are leader groups to be tested is desirable.

The experimental designer then is faced with an alternative selection between candidates, the testing of which is associated with three nuisance factors. In each experimental run there are at least five separate situations that stress the attribute to be measured for each candidate. Thus, in effect we have a square replicated five times. This gives (4-1) (5x4-30= 51 degrees of freedom for error.

IV. SOME OF THE EXPERIMENTAL DESIGNS CONSIDERED. Figure 1* is the design of our first experiment where our alternatives were variations in the number of antitank weapons and mortars. Interactions are not ignored in design. If an important interaction is suspected, the interaction is made the subject of a separate experiment. In this particular experiment, the interactions appeared to be of very little interest since a comparison of the average performance with several leader groups over a variety of terrains was desired.

As our experience increases, it is likely that the variables we consider will be changed. In one of our recent experiments with artillery fire, the effects of learning and key personnel were small compared to those of terrain and the candidates. In a similar experiment utilizing less highly specialized troops, the effects of learning were much stronger. In tests by BRL, Aberdeen Proving Ground. CORG, and CDEC, the effect of learning is the most pronounced of any effect in tank--antitank weapons performance.

Each successive design becomes more refined based on the knowledge of the previous experiment. One of our problems is to anticipate the results of experiments not yet undertaken in the preparation of new designs. We would like not to do this but the pressure of our problems and the lead time required for procurement of equipment training of troops and experimental lay-out demands it.

Figure II is a sample of such a design. This particular experiment is not planned for the present but some similar to it are probable. It was considered at one time. The experiment was designed for determining the types of reconnaissance antitank weapons and transportation system attachments which would be required for a force. The assumption made was that a series of experiments had been completed on smaller sized units. Each experimental run was to last a week with a week in between for a partial data analysis and military assessment of performance. During the first week, three reconnaissance alternatives would be examined using leader group A and a selected antitank and transportation system. The main force would be comprised of three basic types of units. On the basis of a quick analysis, the best reconnaissance would be given leader group B to run over a similar problem while group B tested three antitank weapons systems alternatives. Group B would have four basic units. This would continue with the best of the alternatives based on as many previous runs as possible being used. After nine runs a set of three "confirmation runs" would be selected on the basis of analysis and military judgment.

Let us examine what the first week's "run" might look like in this

experiment in Figure III. Here we study many operations in the course of each day, such as assault, defense, etc. In addition, we are concerned with the span of operations of the force. The control of the force and its operational effectiveness may vary not only with the basic number of organizations but with their zone of responsibility.

As the span of operations is changed, we assign the reconnaissance system most likely to be used on that span. Thus R_3 might be of value only on the span of operations, S_3, but R_2 and R_1 might be competitors over the narrower spans. There are also associated tactics and doctrinal concepts of operations with these spans and reconnaissance systems.

This design exemplifies some important areas in experimental design of large scale field experiments. First is the idea of associated tactics, doctrines, and spans of operations with the alternatives. This may appear confusing to the designer, but change of a single characteristic with all others held constant is actually both unrealistic and frequently meaningless. Second is the idea of the maximum use of professional military judgment along with scientific analysis in designing confirmation experiments. Third is the idea of planned experiment using the learning of early runs to improve on likelihood of gathering meaningful data in subsequent runs. Fourth is the idea of high risk - high return experiments. The last idea must predominate in our present situation. The security of an experiment to "examine a trend" or of one assured by the number of replications to be statistically easy to analyze is not one that we are likely to run because of the waste of military advice implied. A complete failure is unlikely because too much useful information is being gathered. On the other hand, the foregoing experimental design has many possible things that could go wrong.

Not all of our experimental designs are so complicated or fraught with danger. In our small side experiments where the effort an an additional run is low, we build up large sample sizes. In a recent experiment with an infantry hand fired weapon we fired 1,200 rounds of ammunition. Here though, a run took less than fifteen minutes and occupied less than fifteen people.

V. SUMMARY. The experimental design problems of a unique facility of the US Army have been outlined. These problems are of a nature that demand inquiry at a fundamental level of design reasoning following these steps.

A. The criterion of casualties and time which requires two-sided simulated combat.

B. The type of experiment which is discriminatory among candidates.

C. The basis of candidate selection which requires a high level of military judgment to guarantee that the candidates are not only competitive but truly different.

D. The control of the nuisance variables which are the leader group, learning, and terrain variables.

E. The selection of a design whose number of replications is held to a minimum, and where interactions are not usually considered.

The importance of the interplay of scientific and military judgment in the design and analysis of these experiments has been emphasized. A discussion has been made of the influence of the urgency and magnitude of our task on the types of designs that we use. Our work, we feel, is new and demands varied approaches. Eventually, with experience, we expect to simplify our criteria, shorten our experiments, and identify more precisely those variables which we must consider simultaneously.

FIGURE I

EXPERIMENTAL DESIGN FOR SELECTION OF ALTERNATIVE WEAPONS SYSTEMS FOR COMPANY SIZED FORCES

Alternatives

PHASE Training Level		A_1	A_2	A_3	A_4
	I	L_1S_1	L_2S_2	L_3S_3	L_4S_4
	II	L_2S_3	L_1S_4	L_4S_1	L_3S_2
	III	L_3S_4	L_4S_3	L_1S_2	L_2S_1
	IV	L_4S_2	L_3S_1	L_2S_4	L_1S_3

L_1, L_2, etc. are Leader-men groups of company size.

S_1, S_2, etc. are situations with varying Terrain including:

Defense and Attack of a Prepared Position

Defense and Attack of a Delaying Position

FIGURE II

SUGGESTED EXPERIMENTAL DESIGN

and Associated Time Schedule

Phase / No. of units	3	4	5
I	$AR(F_1M_1)$ wk. 1	$BF(R_bM_1)$ wk. 3	$CM(R_bF_b)$ wk. 5
II	$CF(R_bM_b)$ wk. 7	$AM(F^*_bR_b)$ wk. 9	$BR(M^*_bF^*_b)$ wk. 11
III	$BM(R^*_bF^*_b)$ wk. 13	$CR(M^{**}_bF^*_b)$ wk. 15	$AF(M^{**}_bR^{**}_b)$ wk. 17

Order to be determined from Phase III

IV	A wk. 19	B wk. 21	C wk. 23

for	Leader Groups	A B C
	Recon. and Surveil. Alt.	$(R_1R_2$ etc.$) = R$
	A-T Weap. Alt.	$(F_1F_2$ etc.$) = F$
	Trans. Sys. Alt.	$(M_1M_2$ etc.$) = M$
	Best of Alt. Studied	$R_bF_bM_b$

* Based on two runs

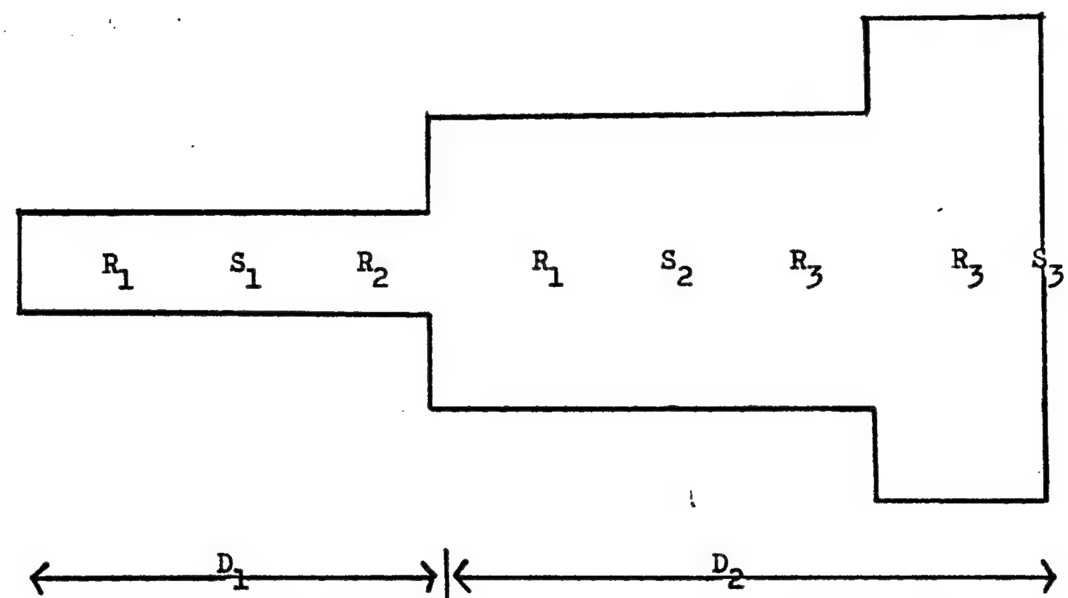
** Based on three runs

FIGURE III

SUGGESTED OPERATIONS FOR A GIVEN RUN

With Leadership A, AT Weapons System F_1, and Mobility System M_1

MON	TUE	WED	THU	FRI
$T_1T_2T_3$	$T_1T_2T_4$	$T_5T_6T_3$	$T_5T_6T_7$	$T_5T_6T_7$



for Span of Operations------------$S_1S_2S_3$

Recon and Surveil Alt---------$R_1R_2R_3$

Tactical Concept--------------D_1D_2

Tactics and Techniques--------$T_1 \ldots T_7$

A POINT OF VIEW IN THE ANALYSIS OF SIMULATION DATA

Sol Haberman
Operations Research Office
Johns Hopkins University

SUMMARY. Instead of visualizing variables as related amond themselves in the form of equations such as:

$$Y = K_1X_1 + K_2K_2 + \dots + K_nX_n \tag{1}$$

or

$$C_1Y_1 + C_2Y_2 + \dots + C_mY_m = K_1X_1 + K_2X_2 + \dots + K_nX_n \tag{2}$$

where the X's are situation variables and the Y's are outcome variables (criteria), it is proposed here that another tack be taken altogether.

It is suggested that each variable be stated in terms of a very few categories and that a set of variables be stated as a set of combinations of categories

It will be demonstrated from a particular set of data that:

1. When variables are treated grossly but combinatorially their relative weights can be measured in a probabilistic sense.
2. There exist at least two methods of analysis which give substantially similar answers.
3. The computations for Methods I and II (which will be explained) are simple.
4. The interrelationships of the variables may be directly inspected in a Table.

It will not be demonstrated here how the relevant probability distributio are derived.

PART I discusses the case of a single criterion.

PART II discusses the technique of handling several criteria.

INTRODUCTION. The point of view explained here in working with complex systems of variables, such as those which are met in simulations, is that the analysis loses little and gains much if we break up the range of each variable into very few points or intervals, preferably two or three. It will be claimed that such an approach does some violence to the data but examples will be given that demonstrate payoff both in clarity of results and in ease of calculations.

PART I

If a variable is thought of as a collection of mutually exclusive slots which carry numerical or verbal labels, a set of variables may be dealt with as a collection of combinations of these slots. For example, in TATS*which

* TATS was prepared by Paul Newcomb and Bernard Urban. The data were obtained by Leon Feldman.

is a simulation where five tanks attack a position held by three anti-tanks, three anti-tank characteristics can be expressed as three variables with two categories (slots) in each,

	P_K	(Probability of Kill)	(+ High) (- Low)
	M_{FT}	(Mean Fire Time)	(L Long) (S Short)
and	T_H	(Time anti-tank remains hidden)	(L Long) (S Short)

TABLE 1

	P_K	M_{FT}	T_H		Ratio - Tanks Killed * to Anti-tanks Killed		
1.	High	Short	Short	(+ SS)	149/96	=	1.55
2.	High	Short	Long	(+ SL)	150/99	=	1.52
3.	High	Long	Long	(+ LL)	135/91	=	1.48
4.	High	Long	Short	(+ LS)	137/99	=	1.38
5.	Low	Short	Short	(- SS)	65/111	=	0.58
6.	Low	Long	Short	(- LS)	63/112	=	0.56
7.	Low	Long	Long	(- LL)	60/116	=	0.51
8.	Low	Short	Long	(- SL)	56/115	=	0.49

* From Forty games for each combination

First Step

First we might ask if the variables point in the same direction in the sense of vectors. If they do, they belong together in the sense that they correlate with the same phenomenon (in this case, the ratio of tanks to anti-tanks killed). We place our combinations geometrically,

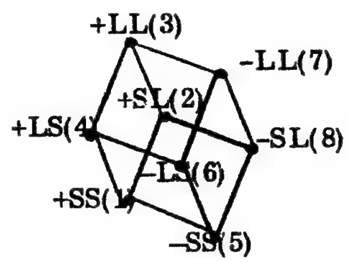


Figure 1.

and find that the ranks associated with them are not randomly distributed and that a vector can be visualized from (+ SS) (1) to approximately (- LL) (7) without undue strain. A test is available, (in Biometrika, Vol. 42, Table 2, P. 421) which tells us that this above distribution of ranks can be achieved or bettered only 1.3 percent of the time by chance alone. We conclude then, that the variables do belong together.

Second Step (Alternatives Illustrated)

Since our variables relate to the criterion we now ask what the relative importance of each is in the set.

Method I. Holding two variables constant and taking differences between ranks we get: *

TABLE 2a

P_K allowed to vary		M_{FT} allowed to vary		T_H allowed to vary	
+ LL - LL	(7-3) = 4	+ LL + SL	(3-2) = 1	+ LL + LS	(4-3) = 1
+ SL - SL	(8-2) = 6	+ LS + SS	(4-1) = 3	+ SL + SS	(2-1) = 1
+ LS - SL	(6-4) = 2	- LL - SL	(8-7) = 1	- LL - LS	(7-6) = 1
+ SS - SS	(5-1) = 4	- LS - SS	(6-5) = 1	- SL - SS	(8-5) = 3

The sums of squares of differences are:

$$P_K = 72, \ (16 + 36 + 4 + 16)$$

$$M_{FT} = 12, \ (1 + 9 + 1 + 1)$$

$$T_H = 12, \ (1 + 1 + 1 + 9)$$

A test would show that P_K outweighs the other two variables significantly. *

It can be seen by referring to the cube that geometrically this is the same set of operations as taking differences along the edges of the cube.

Holding only one variable constant and taking differences between ranks, we get Table 2b. **

The sums of squares of differences are:

$$M_{FT}, T_H = 16 \ (4 + 4 + 4 + 4)$$

$$P_K, M_{FT} = 76 \ (25 + 25 + 1 + 25)$$

$$P_K, T_H = 76 \ (9 + 9 + 9 + 49)$$

* The probability distribution for the test is in the process of computation

** Geometrically this is the same as taking differences across the diagonals.

TABLE 2b

M_{FT}, T_H vary and P_K constant		P_K, M_{FT} vary and T_H constant		P_K, T_H vary and M_{FT} constant	
+ LL + SS	(3-1) = 2	+ LL - SL	(8-3) = 5	+ LL - LS	(6-3) = 3
+ LS + SL	(4-2) = 2	+ SL - LL	(7-2) = 5	+ LS - LL	(7-4) = 3
- LS - SL	(8-6) = 2	+ LS - SS	(5-4) = 1	+ SL - SS	(5-2) = 3
- LL - SS	(7-5) = 2	+ SS - LS	(6-1) = 5	+ SS - SL	(8-1) = 7

A test would show that P_K together with either of the other two variables is significantly more important than the combined value of M_{FT} and T_H as a pair.

Method II. Method II can most be likened to a correlation technique. Its computations are even more simple than those of Method I. If we examine the columns of Table I we see that the column which refers to P_K is as perfectly stated as possible for maximum correlation with the criterion, which is the ratio of tanks to anti-tanks killed. All +'s precede all -'s in the form + + + + - - - - as we look down the column. If we examine the column which refers to M_{FT} we see that the ordering is S S L L S L L S. If would take a minimum of six interchanges of adjacent pairs of letters to correct this disarray so that it would look like S S S S L L L L. Similarly, referring to T_H it would take a minimum of six interchanges of adjacent pairs of letters to straighten out S L L S S S L L.

The number of interchanges is taken as a measure of the importance of a variable, the smaller the number the greater the importance, (in contrast to Method I which is constructed so that the larger the number the greater the importance).

In summary, $P_K = 0$, $M_{FT} = 6$, $T_H = 6$, which corroborates the results obtained using Method I.

If we consider the four possible combinations of P_K and M_{FT} we can again judge the ranking of these with respect to their concentration or grouping in a pair of columns. These are, (+ S), (+ L), (- S) and (- L) and if refer to columns P_K and M_{FT} of Table 1 we see:

+ S
+ S
+ L
+ L
- S
- L
- L
- S

The judgment of the degree of disarray is made, as with single variables, by counting the minimum number of interchanges of adjacent pairs which are needed to eliminate it. In this case, it can be seen visually that if we raise (-S) from last to sixth place we would have:

+ S
+ S
+ L
+ L
- S
- S
- L
- L

a ranking which takes two interchanges to accomplish and which exhibits maximum correlation with the criterion.

Using similar calculations it would take two interchanges to correct the ranking formed from the combinations of P_K and T_H and it would take nine interchanges to correct the ranking formed from the combinations of M_{FT} and T_H.

In summary, * $P_K, M_{FT} = 2$, $P_K, T_H = 2$, and $M_{FT}, T_H = 9$, which again corroborates the results under Method I.

Summary and Future Plans

Work is being done on the probability distributions for Methods I and II and it is expected that some results will be published.

Method I appears to be more sensitive to real differences and is more in line with orthodox notions of experimentation. The probability distributions for Method II, however, seem to be more easily derived. As we have seen, the results for Methods I and II coincide extremely closely for the above data. **

PART II

Introduction

The variables which enter into a simulation may be defined as belonging to either of two classes, either those composing the situation or those describing the outcome. I would categorize the tank and anti-tank characteristics as situation variables and the single criterion against which they were ranked, the ratio of tanks to anti-tanks killed in a series of battles, as the outcome variable.

But real life requires that we measure the relative importance of situation variables not against one criterion but against several simultaneously. A method exists for doing so and TATS data will be used to illustrate this.

*

a. One Variable allowed to vary	Method I	Method II	b. Two Variables Allowed to vary	Method I	Method II
P_K	72	0	P_K, M_{FT}	76	2
M_{FT}	12	6	P_K, T_H	76	2
T_H	12	6	M_{FT}, T_H	16	9

** See Appendices I and II

Since the simulation consists of an attack by five tanks on a position held by three anti-tanks, we can consider the outcomes as combinations of three distinct variables, "tanks killed" (0 to 5), "anti-tanks killed" (0 to 3) and "position taken or not taken" (+ or -).

Example I

Forty games were played for each of the eight combinations listed above. Seven outcome combinations contained 282 of 320 outcomes while 15 outcome combinations contained the remaining 38 outcomes. Omitting the columns which contained fewer than ten outcomes, the original data as copied from the print-outs gave us a matrix containing eight rows and seven columns as follows:

TABLE 3

			Outcomes*							
P_K	M_{FT}	T_H	23+	13+	52-	33+	51-	43+	03+	Total
+	L	L	9	1	8	10	3	1	-	32
-	L	L	13	16	-	3	-	-	4	36
+	S	L	7	1	15	8	1	4	-	36
+	L	S	7	3	8	8	4	6	-	36
-	S	L	10	17	-	3	-	2	4	36
+	S	S	8	1	15	9	2	2	-	37
-	L	L	9	14	1	5	1	-	5	35
-	S	S	13	7	-	4	1	2	7	34
		Total	76	60	47	50	12	17	20	282

* 23+ means two tanks and three anti-tanks killed, and position taken
52- means five tanks and two anti-tanks killed, and position not taken

Now the assumption is, that if we maximize the correlation between the situation variables P, M and T and the outcomes, we are reflecting more closely their relationship in nature.

When this has been done, the ranking of the situation combinations and the ranking of the outcome combinations may be examined for significance. This is another way of saying that the relative importance of the variables may then be stated.

If we define maximum correlation between situation and outcome variables as that permutation of rows and columns which gives a maximum positive sum for all the possible * 2x2 determinants, (in this case there are $\frac{8!}{2!\,6!} \times \frac{7!}{2!\,5!} =$ 28 x 21 such determinants) we get a new matrix:

TABLE 4

P_K	M_{FT}	T_H	Outcomes 13+	03+	23+	33+	51-	43+	52-	Total
+	S	L	1	-	7	8	1	4	15	36
+	S	S	1	-	8	9	2	2	15	37
+	L	S	3	-	7	8	4	6	8	36
+	L	L	1	-	9	10	3	1	8	32
-	S	S	7	7	13	4	1	2	-	34
-	L	S	14	5	9	5	1	-	1	35
-	S	L	17	4	10	3	-	2	-	36
-	L	L	16	4	13	3	-	-	-	36
Total			60	20	76	50	12	17	47	282

Inspecting the matrix of Table 4 we see that the frequencies now appear to be concentrated along the diagonal from the top right corner to the bottom left corner.

If we write the two rankings of the eight situation combinations we have obtained from Tables 1 and 4 side by side we see a tremendously significant correlation ($\rho = +.93$) between them.

TABLE 5

Ratio $\left(\frac{\text{Tanks Killed *}}{\text{Anti-tanks Killed}}\right)$		Determinant Method	
1.	+ SS	2.	+ SL
2.	+ SL	1.	+ SS
3.	+ LL	4.	+ LS
4.	+ LS	3.	+ LL
5.	- SS	5.	- SS
6.	- LS	6.	- LS
7.	- LL	8.	- SL
8.	- SL	7.	- LL

* Shown geometrically in Figure 1 of Part I

* A "possible" determinant is formed from paired rows and columns by the rule $\begin{vmatrix} Mpq & Mpr \\ Msq & Msr \end{vmatrix}$ where p and s are rows, q and r are columns, and M is a frequency no.

If we write the ranking of outcomes twice, once as expected on an apriori basis and once as it actually occured after permuting rows and columns we again see a very strong correlation (ρ = .86) between both.

TABLE 6

A Priori	Determinant Method
1. 03+	2. 13+
2. 13+	1. 03+
3. 23+	3. 23+
4. 33+	4. 33+
5. 43+	7. 51-
6. 52-	5. 43+
7. 51-	6. 52-

Example II

The tank anti-tank simulator was run again on a different series of situation controls playing ten games for each situation. The variables this time were:

(Probability of Kill by the Anti-tanks) (+ High / - Low)

(Mean Fire Time of Anti-tanks) (+ Long / - Short)

(Mean Fire Time of Tanks) (+ Long / - Short)

(Probability of Kill by the Tanks) (+ High / - Low)

When scored as before on the ratio of tanks to anti-tanks killed for each set of ten games, the ranking of the combinations was found to be:

TABLE 7*

	Anti-Tanks		Tanks	
	P_K	M_{FT}	M_{FT}	P_K
1.	+	S	S	–
2.	+	S	L	–
3.	+	L	L	–
4.	+	L	S	–
5.	+	S	L	+
6.	+	L	L	+
7.	–	S	L	–
8.	+	S	S	+
9.	+	L	S	+
10.	–	S	S	–
11.	–	S	L	+
12.	–	L	L	–
13.	–	S	S	+
14.	–	L	L	+
15.	–	L	S	–
16.	–	L	S	+

* The calculations of Table A1, App. I are based on this ranking.

Omitting columns with negligible frequencies, the table obtained from the print-outs was:

TABLE 8

P_K	M_{FT}	M_{FT}	P_K	Outcomes 52–	33+	51–	23+	43+	13+	03+	Totals
+	L	L	+	2	3	–	5	–	–	–	10
+	L	L	–	5	–	2	2	–	–	–	9
+	L	S	–	4	–	3	3	–	–	–	10
+	S	S	–	4	1	3	–	2	–	–	10
–	S	S	–	–	1	–	3	–	4	–	8
–	S	S	+	–	1	–	4	–	3	2	10
–	S	L	+	1	5	–	–	–	1	3	10
–	L	L	+	–	–	–	4	–	3	3	10
–	L	L	–	–	1	–	2	–	3	2	8
–	L	S	–	–	1	–	1	–	1	5	8
–	L	S	+	–	–	–	–	–	4	6	10

(Table 8 Continued

TABLE 8 (Continued)

P_K	M_{FT}	M_{FT}	P_K	52-	33+	51-	23+	43+	13+	03+	Totals
+	L	S	+	-	1	-	8	-	-	-	9
+	S	S	+	1	2	-	3	2	2	-	10
+	S	L	+	-	3	-	2	4	-	-	9
+	S	L	-	2	1	3	1	2	-	-	9
-	S	L	-	-	1	1	1	1	4	-	8
			Total	19	21	12	39	11	25	21	148

Permuting rows and columns so that the positive sum of the determinant values was as close to a maximum as possible, Table 9 was obtained:

TABLE 9

				Outcomes							
P_K	M_{FT}	M_{FT}	P_K	03+	13+	33+	23+	43+	51-	52-	Totals
+	L	L	-	-	-	-	2	-	2	5	9
+	S	S	-	-	-	1	-	2	3	4	10
+	L	S	-	-	-	-	3	-	3	4	10
+	S	L	-	-	-	1	1	2	3	2	9
+	S	L	+	-	-	3	2	4	-	-	9
+	L	S	+	-	-	1	8	-	-	-	9
+	L	L	+	-	-	3	5	-	-	2	10
+	S	S	+	-	2	2	3	2	-	1	10
-	S	L	-	-	4	1	1	1	1	-	8
-	S	S	-	-	4	1	3	-	-	-	8
-	S	S	+	2	3	1	4	-	-	-	10
-	L	L	+	3	3	-	4	-	-	-	10
-	S	L	+	3	1	5	-	-	-	1	10
-	L	L	-	2	3	1	2	-	-	-	8
-	L	S	-	5	1	1	1	-	-	-	8
-	L	S	+	6	4	-	-	-	-	-	10
			Total	21	25	21	39	11	12	19	148

If we write the two rankings obtained by the two methods for the situation variables side by side, we see again a very great correlation between the two arrays of the order of $\rho = +.94$.

TABLE 10

	Ratios (from T. 7)					Method of Determinants (Vertical axis of T. 9)			
	P_K	M_{FT}	M_{FT}	P_K		P_K	M_{FT}	M_{FT}	P_K
1.	+	S	S	-	3.	+	L	L	-
2.	+	S	L	-	1.	+	S	S	-
3.	+	L	L	-	4.	+	L	S	-
4.	+	L	S	-	2.	+	S	L	-
5.	+	S	L	+	5.	+	S	L	+
6.	+	L	L	+	9.	+	L	S	+
7.	-	S	L	-	6.	+	L	L	+
8.	+	S	S	+	8.	+	S	S	+
9.	+	L	S	+	7.	-	S	L	-
10.	-	S	S	-	10.	-	S	S	-
11.	-	S	L	+	13.	-	S	S	+
12.	-	L	L	-	14.	-	L	L	+
13.	-	S	S	+	11.	-	S	L	+
14.	-	L	L	+	12.	-	L	L	-
15.	-	L	S	-	15.	-	L	S	-
16.	-	L	S	+	16.	-	L	S	+

Also, the a priori ranking of the outcome variables correlates with the obtained ranking, ($\rho = +.91$).

Thus, after permuting rows and columns without reference to row and column labels, according to the requirement that all the possible 2 x 2 determinants give the maximum positive sum we have achieved two rankings which coincide with what we know or expect. The assumption that if we maximize the correlation between situation and outcome variables we thereby reflect more closely their relationship in nature seems to be confirmed by the examples.

There is not as yet a method which can be stated in purely mathematical form for finding the optimum permutation of rows and columns nor can the class of matrices to which this method applies be exactly stated. If there is some "scatter" in each row and column ("scatter," being defined at this stage as

consisting of at least two non-zero frequencies), the "best" rankings of the row and column labels can be found. Examining the non-stochastically distributed matrix of the form

		M_{13}
	M_{22}	
M_{31}		

the method of determinants cannot distinguish it from its permutation

		M_{31}
	M_{13}	
M_{22}		

An interesting matrix is:

Set x \ Set y				
	1	2	3	4
	5	6	7	8
	9	10	11	12
	13	14	15	16

which exhibits maximum correlation in its stated order of rows and columns and which shows that the method of determinants need not and cannot always concentrate values toward the diagonal. Situations can arise as above where we get simultaneous increases in frequencies as we move along each of the scales and along the diagonals. These situations might be viewed in the light of a more liberal definition of what is meant by correlation.

To illustrate the mechanics of what was done in the 8 x 7 matrix, let us work with a smaller matrix by consolidating the table into a 4 x 7 matrix by omitting consideration of variable. T_H.

We now have Table 11 and we wish to maximize the correlation between the two sets of variables:

TABLE 11

	Situations		Outcomes (1)	(2)	(3)	(4)	(5)	(6)	(7)	
	P_K	M_{FT}	23+	13+	52−	33+	51−	43+	03+	Totals
(1)	+	L	16	4	16	18	7	7	−	68
(2)	−	L	22	30	1	8	1	−	9	71
(3)	+	S	15	2	30	17	3	6	−	73
(4)	−	S	23	24	−	7	1	4	11	70
		Total	76	60	47	50	12	17	20	282

We form a new matrix, the elements of which are the $(\frac{4!}{2!\,2!} \times \frac{7!}{2!\,5!})$ or (6 x 21) 2 x 2 determinants. The columns of this derived matrix are labeled according to which pair of columns of the basic matrix are being considered and similarly the new rows are labeled by pairs of original rows, as follows: (See Table 12)

Keeping the plus and minus signs in the body of the matrix and examining the marginal sums in terms of + and − we get: (See Tables 13, 14, 15)

The row sums cannot be improved by any further alterations in sign so we stop at this point. (With the 16 x 7 matrix presented before, it was found necessary to go up and back to row and column sums many more times than was done here to arrive at the maximum positive sum.) If we examine the logical consequences for columns by saying 2 precedes 1, 1 precedes 3, 4 precedes 1 etc., as we have written them along the Y axis of Table 15, we get that $2 > 4 > 1 > 3$ (where > means precedes), and similarly the logical consequences for columns are $7 > 2 > 1 > 4 > 5 > 6 > 3$.

TABLE 12

	1, 2		1, 3		
1, 2	16	4	16	16	. . .
	22	30	22	1	
		(+392)		(-336)	
1, 3	16	4	16	16	. . .
	15	2	15	30	
		(-28)		(+240)	
1, 4	16	4	16	16	. . .
	23	24	23	-	
		(+292)		(-368)	
2, 3	22	30	22	1	. . .
	15	2	15	30	
		(-406)		(+645)	
2, 4	22	30	22	1	. . .
	23	24	23	-	
		(-62)		(-23)	
3, 4	15	2	15	30	. . .
	23	24	23	-	
		(+334)		(-690)	
Sums	+1018		+ 885		
	- 496		-1417		

17

TABLE 13

Pairs of Rows	Pairs of Columns 1,2	1,3	1,4	1,5	1,6	1,7	2,3	2,4	2,5	2,6	2,7	3,4	3,5	3,6	3,7	4,5	4,6	4,7	5,6	5,7	6,7	Sums of Values of determinants for rows	
1,2	+	-	-	-	-	+	-	-	-	-	+	+	+	-	+	-	-	+	-	+	+	1,123 - 2,404	3,527
1,3	-	+	+	-	-	0	+	+	-	+	0	-	-	-	0	-	-	0	+	0	-	993 - 716	1,709
1,4	+	-	-	-	-	+	-	-	-	-	+	+	+	+	+	+	+	+	+	+	+	1,276 - 2,051	3,327
2,3	-	+	+	+	+	-	+	+	+	+	-	-	-	+	-	+	+	-	+	-	-	2,839 - 1,603	4,442
2,4	+	-	-	-	+	+	-	+	+	+	+	+	+	+	+	+	+	+	+	+	-	468 - 180	648
3,4	+	-	-	-	-	+	-	-	-	+	+	+	+	+	+	-	+	+	+	+	+	1,665 - 2,296	3,961
																						8,364 - 9,250	17,614
Sums of Values of Determinants for Columns	+1018	+885	+256	+51	+220	+520	+1586	+544	+94	+446	+216	+439	+56	+194	+661	+38	+129	+572	+58	+175	+206	8,364	
	-496	-1417	-890	-395	-338	-135	-1604	-1310	-442	-362	-18	-781	-189	-121	-270	-138	-67	-153	-7	-27	-90	-9,250	
	1514	2302	1146	446	558	655	3190	1854	536	808	234	1220	245	315	931	176	196	725	65	202	296	17,614	

Working first with sums of rows because it seems more convenient, we see that rows (1,2) (1,4) and (3,4) should be alternated in sign. This is the logical equivalent of changing their relative positions in Table 11. This set of alternations in sign affects the column sums. Table 14 gives us the new table:

TABLE 14

Pairs of Rows	Pairs of Columns 1,2	1,3	1,4	1,5	1,6	1,7	2,3	2,4	2,5	2,6	2,7	3,4	3,5	3,6	3,7	4,5	4,6	4,7	5,6	5,7	6,7	Sums of Values of Determinants for Rows	
2,1	–	+	+	+	+	–	+	+	+	+	–	–	–	+	–	+	+	–	+	–	–	2,404 – 1,123	3,527
1,3	–	+	+	–	–	0	+	+	–	+	0	–	–	–	0	–	–	0	+	0	–	993 – 716	1,709
4,1	–	+	+	+	+	–	+	+	+	+	–	–	–	–	–	+	–	–	–	–	–	2,051 – 1,276	3,327
2,3	–	+	+	+	+	–	+	+	+	+	–	–	–	+	–	+	+	–	+	–	–	2,839 – 1,603	4,442
2,4	–	–	–	–	+	+	–	+	+	+	+	+	+	+	+	+	+	+	+	+	–	468 – 180	648
4,3	–	+	+	+	+	–	+	+	+	–	–	–	–	–	–	+	–	–	–	–	–	2,296 – 1,665	3,961
																						11,051 – 6,563	17,614
Sums of Values of Determinants for Columns	0	2279	1112	388	549	35	3166	1854	534	672	114	7	1	17	11	111	136	25	38	2	0	11,051	
	-1514	-23	-34	-58	-9	-620	-24	0	-2	-136	-120	-1213	-244	-298	-920	-65	-60	-700	-27	-200	-296	-6,563	
	1514	2302	1146	446	558	655	3190	1854	536	808	234	1220	245	315	931	176	196	725	65	202	296	17,614	

Next with sums of columns, we see that columns (1, 2), (1, 7), (2, 7), (3, 4), (3, 5), (3, 6), (3, 7), (5, 7), and (6, 7) should be alternated in sign too. As with rows, this is the logical equivalent of reversing their relative positions in Table 11. This set of alternatives now affects the row sums:

TABLE 15

Pairs of Rows	Pairs of Columns 2,1	1,3	1,4	1,5	1,6	7,1	2,3	2,4	2,5	2,6	7,2	4,3	5,3	6,3	7,3	4,5	4,6	4,7	5,6	7,5	7,6		Sums of Values of Determinants for Rows	
2,1	+	+	+	+	+	+	+	+	+	+	+	+	+	−	+	+	+	+	+	+	+		3,520 − 7	3,527
1,3	+	+	+	−	−	0	+	+	−	+	0	+	+	+	0	−	−	+	+	0	+		1,565 − 144	1,709
4,1	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	−	+	−	+	+		3,283 − 44	3,327
2,3	+	+	+	+	+	+	+	+	+	+	+	+	+	−	+	+	+	+	+	+	+		4,436 − 6	4,442
2,4	+	−	−	−	+	−	−	+	+	+	−	−	−	−	−	+	+	+	+	−	+		367 − 281	648
4,3	+	+	+	+	+	+	+	+	+	−	+	+	+	+	+	+	−	+	−	+	+		3,793 − 168	3,961
																							16,964 − 650	17,514
Sums of Value of Determinants for Columns	1514	2279	1112	388	549	620	3166	1854	534	672	120	1213	244	298	920	111	136	700	38	200	296	16,964		
	0	-23	-34	-58	-9	-35	-24	0	-2	-136	-114	-7	-1	-17	-11	-65	-60	-25	-27	-2	0	650		
	1514	2302	1146	446	558	655	3190	1845	536	808	234	1220	245	315	931	176	196	725	65	202	296	17,514		

In none of the calculations so far have logical inconsistencies of the form 1 > 2 > 3 > 1, been obtained and it may be a property of the method that none such can arise. If they do arise, a rule would state that a row or column would be given precedence according to the magnitudes of the sums of determinants associated with it. Using the results obtained, Table 11 is now restated:

TABLE 16

	P_K	M_{FT}	(7) 03+	(2) 13+	(1) 23+	(4) 33+	(5) 51-	(6) 43+	(3) 52-	Totals
(2)	-	L	9	30	22	8	1	-	1	71
(4)	-	S	11	24	23	7	1	4	-	70
(1)	+	L	-	4	16	18	7	7	16	68
(3)	+	S	-	2	15	17	3	6	30	73
		Totals	20	60	76	50	12	17	47	282

By the concentration of values from top left to bottom right it appears we have succeeded in maximizing correlation. Be definition, any vertical or horizontal mirror image of this table is considered the identical table.

The only misplaced column is (51-) and it has the smallest frequency total, 12. Had we obtained more observations, chances are it would have not been misplaced.

The ranking of outcomes correlates with expected ranking with a $\rho = .89$.

The ranking suggests, as before, because of the

P_K	M_{FT}
-	L
-	S
+	L
+	S

- - + + pattern versus the LSLS pattern, that "probability of kill" by the anti-tank is a more important variable than "mean fire time."

3 One variable held constant (three allowed to vary).

Triplets of Variables	ABD	614	(1)	5	(1)
	ACD	606	(2)	8	(2)
	ABC	558	(3)	11	(3)
	BCD	214	(4)	30	(4)

APPENDIX II

Although it is not necessary to the calculations of Method I, Figure 2 gives a visual appreciation of the complexities of the interrelationships measured when dealing with four variables. When one variable is allowed to vary we are taking differences along edges; when two are allowed to vary we are taking differences across the diagonals of the rectangles; and when three are allowed to vary we are taking differences across the diagonals of the cubes.

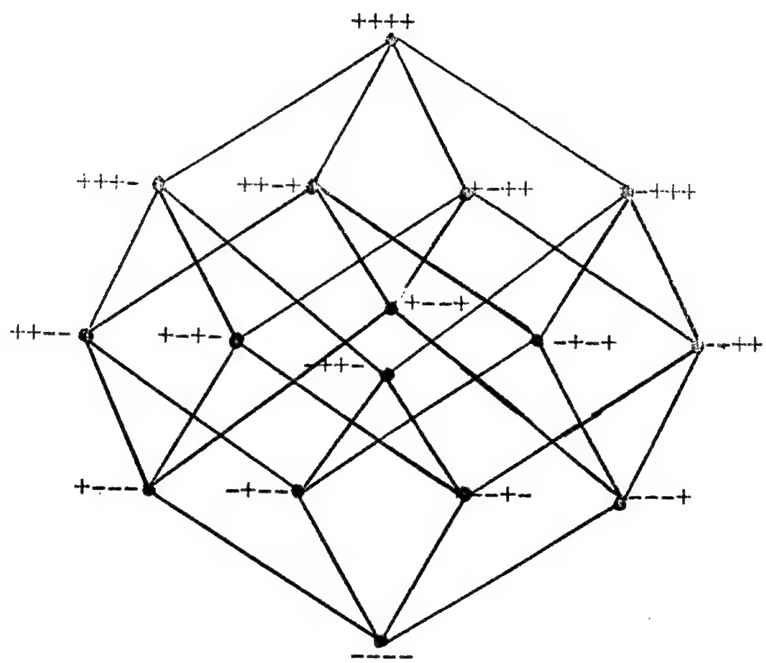


Figure 2

ULTRASONICS, A TOOL FOR WELDMENT INSPECTION

James E. Kingsbury, Wayman N. Clotfelter
and William R. Lucas
Army Ballistic Missile Agency

ABSTRACT. An inspection technique involving the use of low-level untrasonic energy is described. This inspection system was designed for production use in the fabrication of fusion welded pressure vessels. Inspection by this system is done in the shop while the welding fixture is still in place, thus simplifying requried repairs and saving time.

The world of physics is filled today with superlatives. For want of descriptive words to explain the many new developments, some few of the old prefixes are becoming a part of many of the more commonly used words and phrases. One of the most overworked of these prefixes is "ultra." We hear it used in connection with such things as electricity, light and sound. Although Webster defines the prefix as meaning "super or beyond," the use of ultra is made to meet the scientist's needs. One such example of this is the word ultra sound or ultrasonics. The scientific translation of these words is sound above the audible frequency range. It can be anywhere from just barely above the human audible frequency range to infinity. Ultrasonic energy has been employed successfully as an inspection tool in many cases, however due to the many difficult conditions which must be met in an inspection system utilizing ultrasonics, the use of the tool has been limited. Further, due to the bad publicity ultrasonics received from the American Medical Association when its use was first introduced, a scare factor further impeded its development. As is so often the case in new developments, a little knowledge can prove most detrimental. In this case, it was known by AMA that high energy ultrasonics could cause body tissue to deteriorate. The fact not appreciated was that inspection systems utilizing ultrasonci energy were operated at energy levels of less than 1 watt, usually less than ½ watt. Upon clarification of this point, progress increased and today many time and labor consuming inspection systems have been replaced by simple, yet more efficient ultrasonic energy inspection systems. Probably the most attractive use of an ultrasonic inspection system is in the production of welded pressure vessels. The system described in this paper was designed for use with large, fusion welded, pressure vessels as are used in liquid-propelled guided missiles; however, the same general application could be modified for use in small containers.

The need for a reliable, yet simple, system for the quality control of fusion welds in pressure vessels has long been evident. Where the pressure vessel is also a structural component of an assembly, as in the case of some large, liquid-propelled guided missiles, this need becomes critical. To date, the most commonly used inspection tool has been radiography. Although this tool is satisfactory, it leaves much to be desired. For example, there is the problem of transporting the container to a radiation proof laboratory, the time involved in setting up and making the radiographic plates, the developing time and finally the evaluation time. This time is lost as far as the fabricator is concerned. Further, upon detection of defective weldments, the fabricator must replace his

welding fixtures, rout out the defective weldment and repair it, and then the radiographic inspection procedure starts again. It was theorized that low-level ultrasonic energy could be utilized in an inspection system and that the inspection could proceed concurrently with an automatic welding operation. Then, a recording of the ultrasonic scan of the weldment could be evaluated immediately after the weld was completed and necessary repairs made on the original welding set-up, resulting in considerable time saving.

Many problems arose early in the development of an ultrasonic energy inspection system for fusion welds. First, some medium of transmission had to be utilized in getting the energy from a transducer into the weldment. This medium had to be such that it would cause no difficulty where repair welding was necessary. This condition ruled out the use of greases. Water was considered satisfactory but it was not considered feasible to submerge a container of the size in question, in excess of six feet in diameter and of the order of 40 feet long. The use of water jets proved satisfactory and, more recently, the use of a static water column has proven very successful. The latter method reduces considerably the water spillage in the inspection procedure. Secondly, it was necessary to develop a technique for getting as much of the transmitted energy into the weldment as possible. Since the energy level was low, large losses could not be tolerated in transmission. To introduce the ultrasonic energy directly into the crown of the weld bead was ruled out due to the surface roughness. Should the sound be directed at the crown, large and variable reflection losses would occur. Therefore, it was decided to introduce the ultrasonic energy into the parent metal sheet adjacent to the weldment. This presented a smooth surface to the sound beam which reduced the reflection losses. The remaining losses were constant. Since the speed of sound is greater in aluminum than in water, the sound beam was bent when entering the metal, causing it to travel through the sheet, and consequently the weldment, in the desired direction. The controlling factors in deciding the direction which the ultrasonic energy was to travel through the weldment were twofold. First, it was essential that as much of the energy as possible be transmitted through the weldment, the ideal situation having the direction of the energy parallel to the sheet surface. Secondly, since the system utilized a transmission principle rather than reflection, it was essential to provide a means of getting the sound out of the metal. The path chosen is shown in figure 1*. Maximum transmitted energy could be received at either point "A" or "B". Since it is desirable to have both transducers located on the same side of the sheet, the energy beam is allowed to reflect from point "A" and is picked up at point "B". It is not essential that the receiving transducer be located exactly at this point, so long as the location remains constant. It is desirable that the location be close to this point, however, since the transmitted energy is low.

It was further theorized that the welding bar placed behind the joint prior to welding, could be left in place during the inspection. This would be possible since no bond exists between this back-up bar

* Figures are at the end of this article.

and the container being welded. Therefore, an interface is formed between the two pieces which would not allow the energy to leave the material but rather, would cause it to be reflected as mentioned previously. The major advantage here was the capability of the system to inspect the weldment without requiring the welding fixtures to be removed. This would allow immediate repairs to be made on the original welding set-up.

In summary, the theoretical ultrasonic energy inspection system (1) could be employed without endangering personnel in the shop area, (2) could be utlized while the production component was still on the welding jig, (3) would eliminate all processing time in that an instantaneous recording would be made, (4) could be preset to self monitor the weldment, marking on the weldment those areas requiring repair, and (5) would allow necessary repairs to be made simply and without excessive time losses.

The basic theory employed in the inspection system is that ultrasonic energy transmitted through a fixed path will have a constant energy loss. If, however, the path contains voids or discontinuities, additional energy will be lost since these discontinuties create interfaces causing the energy to be reflected out of the path. By recording the energy transmitted, variations in this quantity would indicate defective areas. The energy path is made up of wrought sheet metal and a weldment, thus it is reasonable to assume the discontinuities occur in the weldment. Since the energy measurement is made electronically, it is extremely sensitive. Laboratory testing proved conclusively that the ultrasonic inspection could be made more sensitive than radiographic inspection. However, it also could be preset to indicate a predetermined level of defects and to overlook all defects considered to be of negligible consequence. Testing was initiated in the laboratory with the prime objectives of (1) determining practicability of the theoretical system outlined, (2) determining the reliability and reproducibility of the system, and (3) determining the adaptability of the system to a production operation.

To determine the practicability of the system, a series of defects similar to those commonly found in fusion weldments were machined into a sheet of aluminum. This left no doubt as to the size and number of defects present. The defects were made to simulate such things as linear and transverse weld cracks, random and linear porosity, both large and small, and isolated voids. The instrumentation circuit was initially set up using the Sperry UR reflectoscope with an RA recording and signalling attachment. In conjunction with this, a standard Brush oscillograph was used to record the transmitted energy, thereby producing a permanent record. The transmission link for the ultrasonic energy was a water jet. Initial results indicated one factor had been overlooked in the theoretical determination. The indicated defects on the recording, in all cases, were considerably larger than the actual defect. A simple explanation for this was that the ultrasonic energy beam had a finite width. To take a hypothetical case, let us assume a defect consists of a small void, 1/8" in diameter. As the leading edge of the energy beam arrives at this defect, energy is lost. This energy loss will increase until the defect is located centrally in the beam and then will decrease until the beam's trailing edge leaves the defect. By a series of tests, and careful measurements of the indicated defect and the actual defect, the width of the beam was found to be 5/16". With this value known, it was then possible to determine accurately the defect size. This experimentation also

established that the entire weldment located between the sheet surfaces was being subjected to the energy beam.

In order to determine if linear porosity could be distinguished from linear cracks on the basis of the sound recording, several samples of these defects which were located by radiography were inspected by the ultrasonic system. By linear crack is meant cracks oriented in the direction of the weld bead. Figure 2 which shows a radiograph and a sound recording of the same plate indicates that this differentiation can be made. The testing indicated the presence of linear cracks caused extremely large losses of ultrasonic energy as compared to the loss caused by porosity or voids.

Next, the testing turned entirely to the use of fusion welded plates. A series of plates was prepared and radiographs were made on each. Sound recordings were also made on these plates. From a comparison of the records, it was shown that a good correlation existed. It was determined from these records that by proper calibration of the gail settings, those defects of negligible consequence could be overlooked. This would allow the system to be self-monitoring from a selected presetting. To do this, a small amplifier was built which amplified the signal sent to the recorder. This signal was then clipped electronically so that only those areas where the signal dropped below the selected presetting are indicated as defective. All other areas are indicated at a constant level on the recording. The clipper can be adjusted. By using a small spray type gun which is activated when the signal drops below the preset level, a small dye or paint spray marks the weld area found defective. Concurrently, the marking is also accomplished on the record by first marking 3 foot intervals on the weld and then manually introducing a pip on the recorder at each point. In order that the recording can be matched to the weld, the motor in the oscillograph was replaced with a variable speed motor so the speed of the container being inspected and the speed of the oscillograph are matched, thus producing a 1:1 ratio recording.

One problem, inherent in the system, is the inability to inspect for defects located in either the weld build-up or fall-through. The sound beam coming to the receiving transducer has not come through weld build-up or fall-through. Figure 3 shows sound recordings of a weldment before and after holes were drilled in the crown of the weld. Only two holes, extreme right of figure 3, were drilled deeper into the weld than the top surface of the welded sheet, and these were the only holes detected by the ultrasonic system. The height of the weld crown is usually small compared to the thickness of the welded sheet. A defect not visible at the surface and not extending beneath the top surface of the welded sheet would be considered insignificant and as not requiring repair.

Based on the data accumulated in laboratory tests, the ultrasonic energy inspection system was determined to be practical and results were found to be reproducible. The system used in the laboratory is shown in figure 4. With these fixtures it was possible to move the plate being inspected past the transducers in a manner similar to that which would be encountered in a production set-up. The water jet coupling is shown in detail in figure 5. Although the water jet is practical, it is desirable from a production standpoint to eliminate the water flow. Therefore, the water transmission link was redesigned. To eliminate the water flow, a stagnant water pool is now maintained. This design is

shown in figure 6. To allow angular adjustment of the transducers, a ball and socket joint is incorporated. The water is contained by connecting the sleeve to the transducer on one end and on the other end to the welded container by means of pressure against a rubber seal. A standing column of water assures the pool remains full by supplying make-up water for that lost through seepage as the container moves by the transducers. This system essentially eliminates the water spillage.

A welded container to be inspected by this system must be essentially round so that the critical angles involved do not change seriously and to avoid loss of water from the transmission link. However, this presents no problem because it is equally essential that the container be round for automatic welding. Therefore, the welding back-up bar is machined round and, when in position, it is expanded under pressure thereby assuring the container is round.

The system described has been under development since December 1955. It is presently undergoing extensive calibration in order to allow its use in production. Summarizing the advantages of this system over other non-destructive testing systems which might be used for inspection of fusion-welded pressure vessels, this system (1) is versatile, it can be preset for self-monitoring thereby eliminating the human element in evaluation, (2) does not require special protective measures for personnel, (3) affords tremendous time saving in inspection since an instantaneous recording is produced automatically, and (4) because of its unique capability to make the inspection with the welding fixtures in place, simplifies the repair procedure. Although it should not be construed as a system which will remove a need for radiography, the use of ultrasonic energy as a tool for the inspection of fusion weldments in a production operation offers the several advantages listed not available in any other non-destructive testing method.

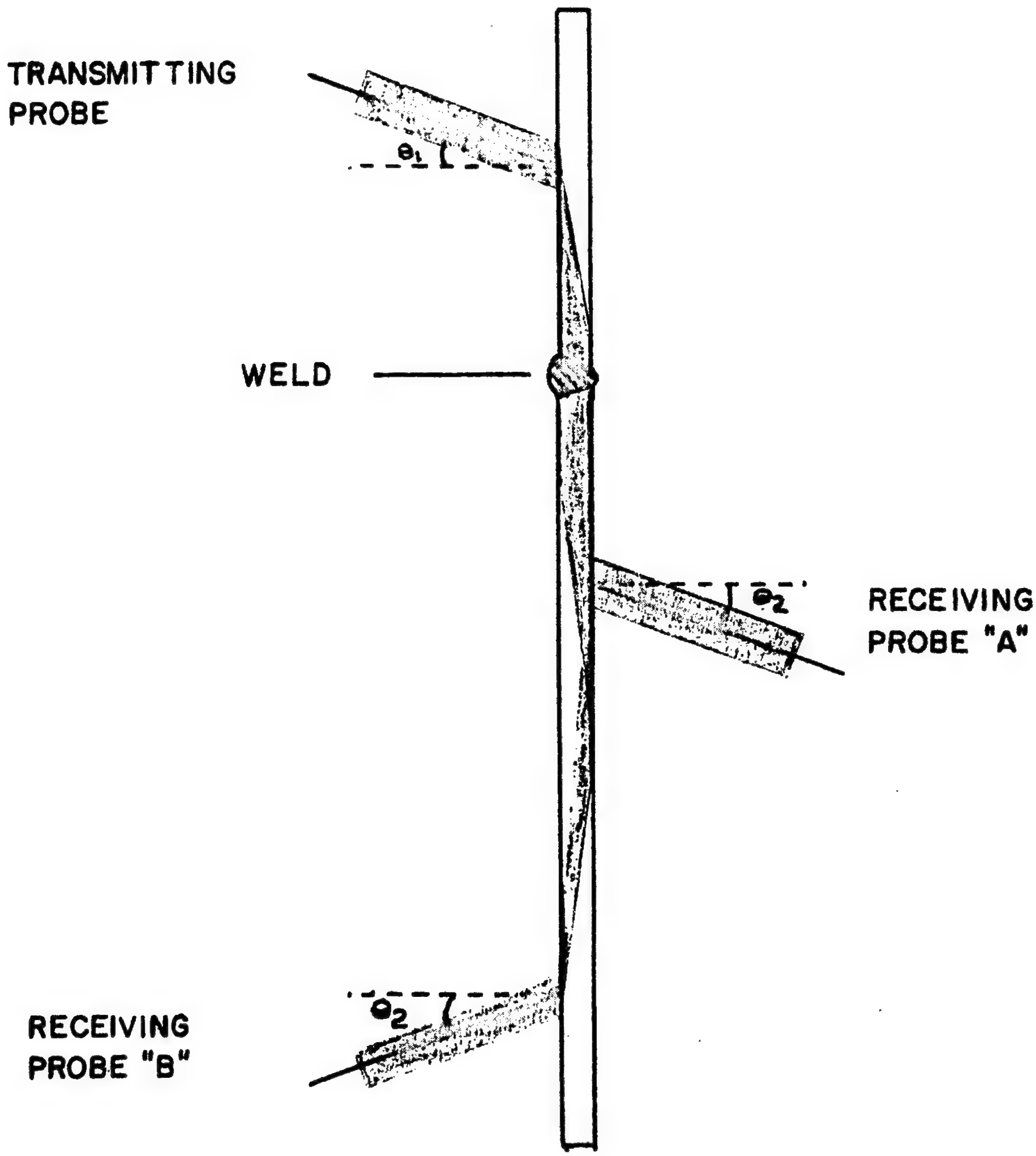
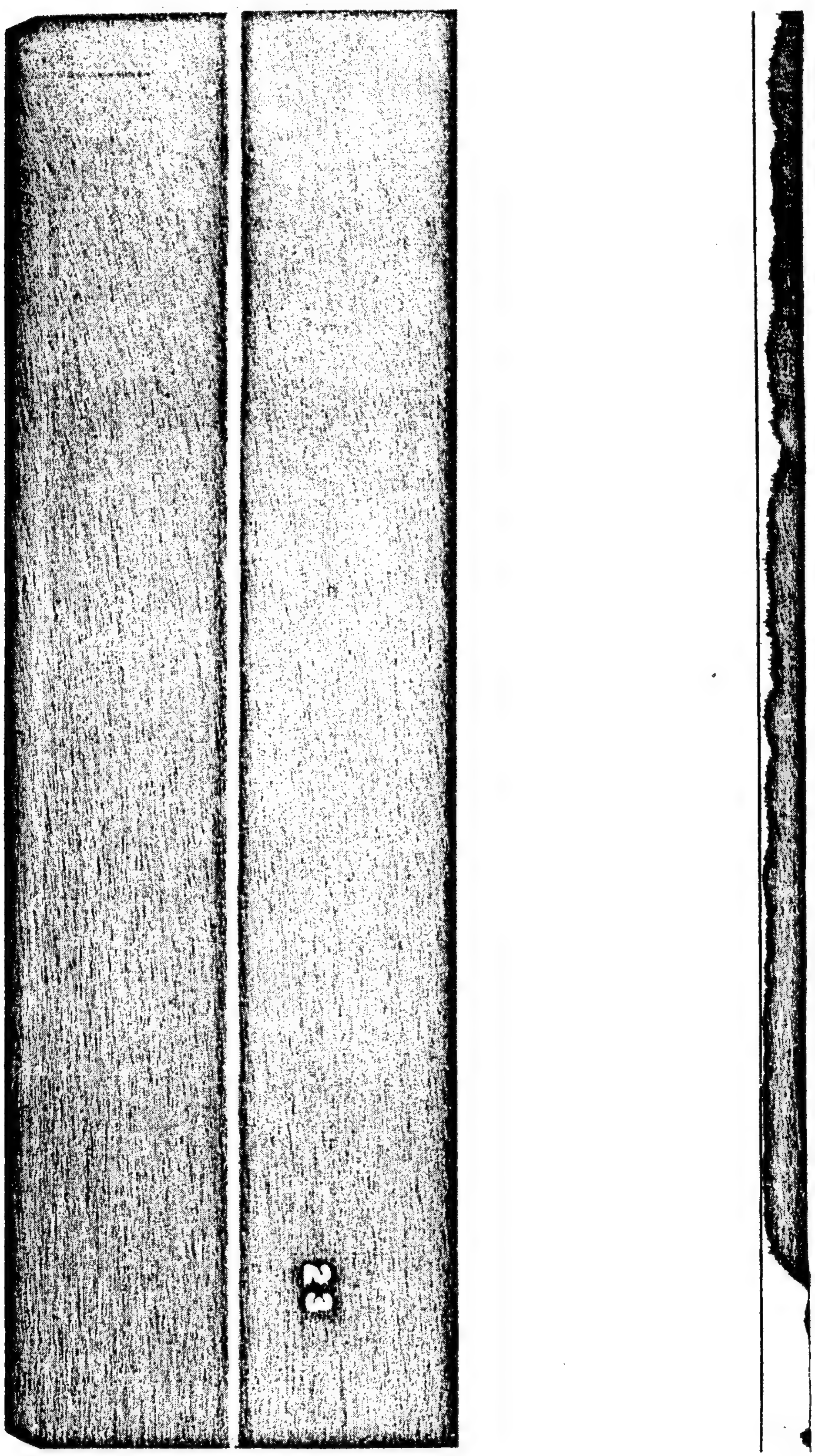


FIG. I PATH OF SOUND THRU WELD

θ_1 IS THE ANGLE OF INCIDENCE

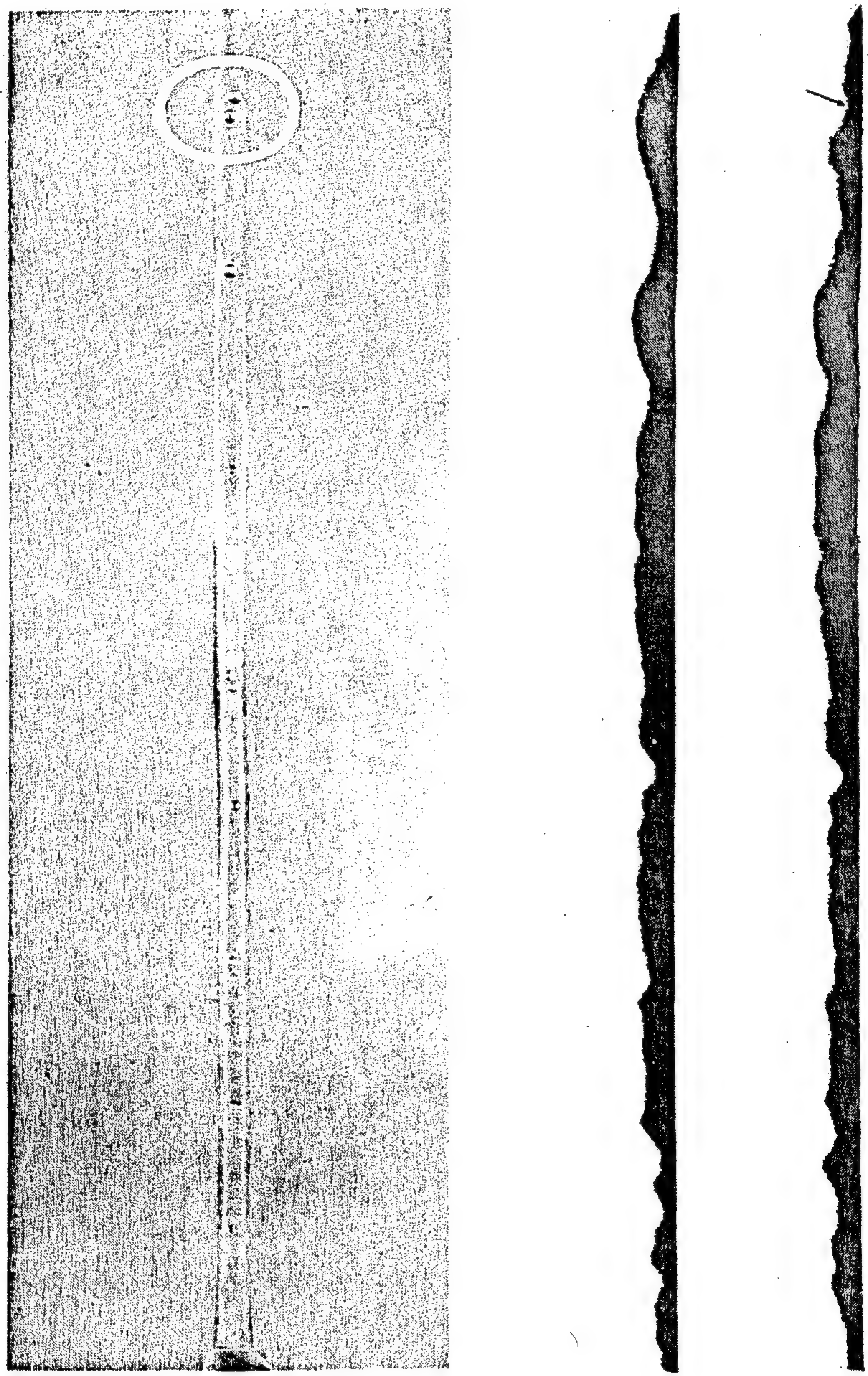
θ_2 IS THE ANGLE OF REFLECTION



(A)

(B)

Fig. [illegible] Comparison of Radiographic and Sonic Inspection Results



(A) (B) (C)

Fig. 3 (A) Weld Plate Containing Defects Located in Weld Build-Up
(B) Sound Recording of Plate Before Holes Were Drilled
(C) Sound Recording of Plate After Holes Were Drilled

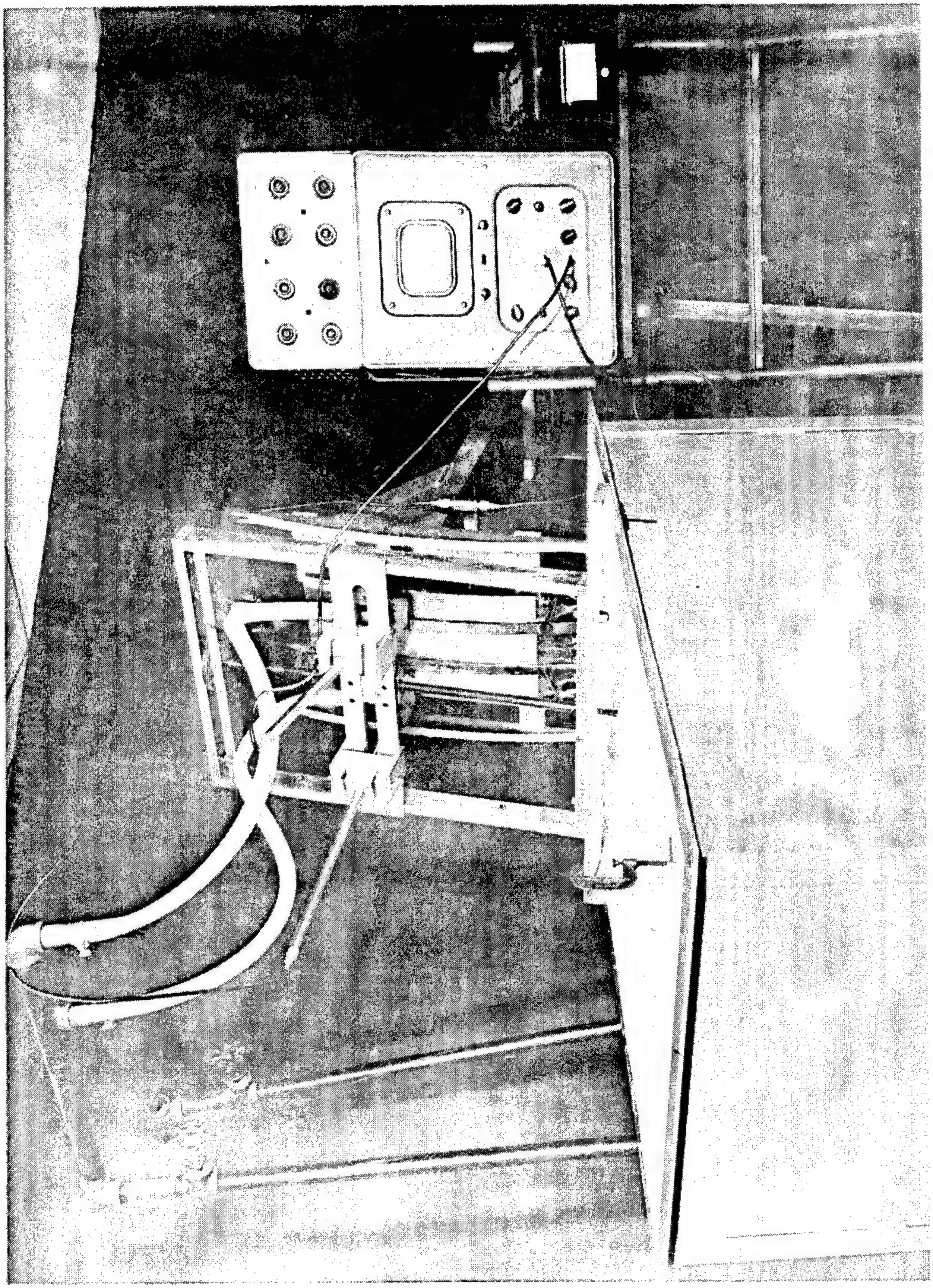


Fig. 5 Water Jet Coupling

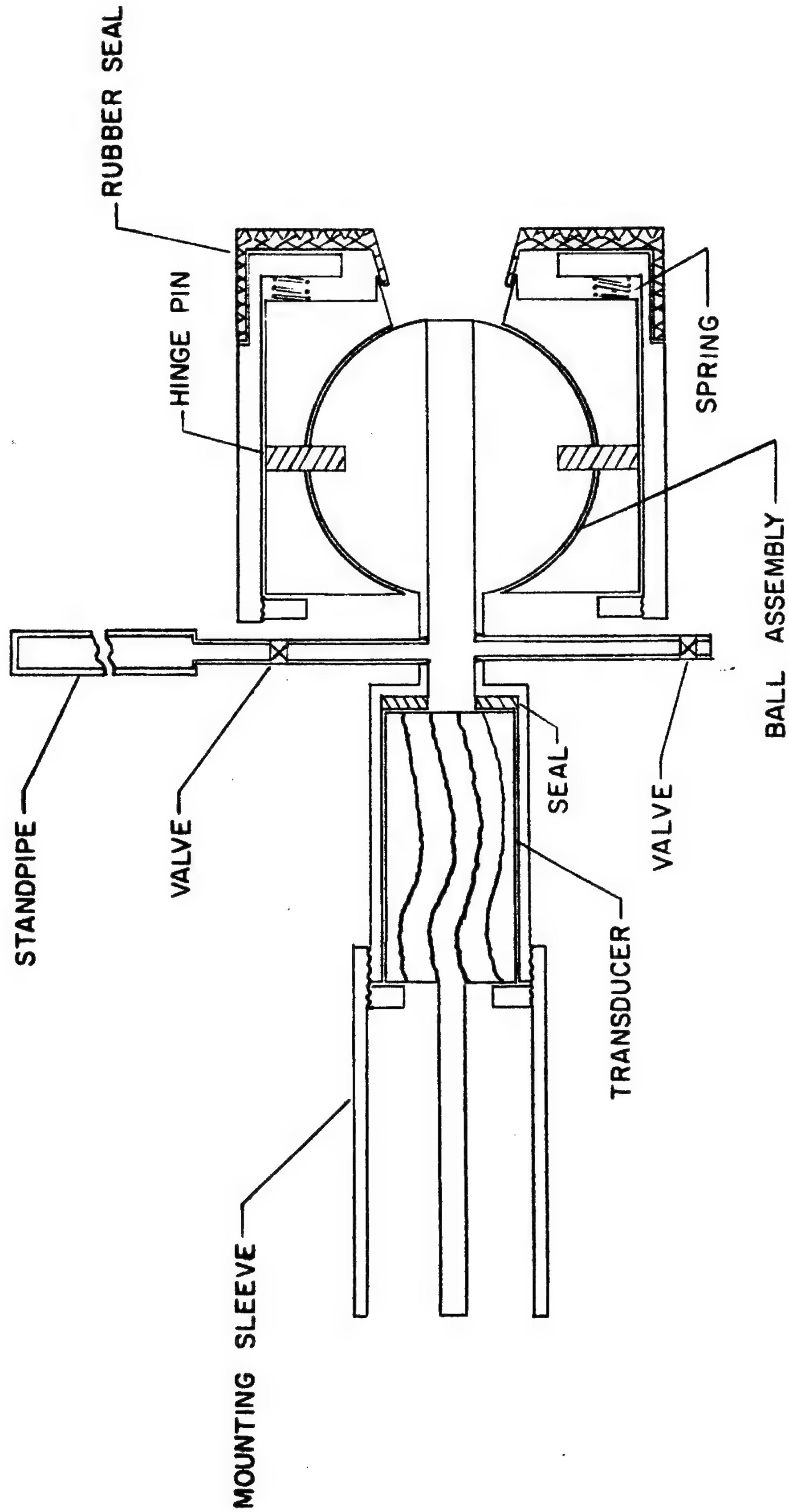


FIG. 6 SKETCH OF STATIC WATER COLUMN COUPLING

SHORT-LIFE STUDY OF CAPACITORS

Robert W. Tucker
Diamond Ordnance Fuze Laboratories

ABSTRACT. To achieve reliability and maximum utilization of space for capacitors in ammunition items such as mortars and missiles, a rapid test was needed to determine the maximum safe voltage to which a capacitor could be subjected when exposed to a wide range of environmental conditions. In this study, impregnated and unimpregnated capacitors having a dielectric of 0.25-mil-thick polyethylene terephthalate were exposed to varied temperatures, and to voltages applied for 1000 seconds. Test data were obtained in such a manner that a voltage-dosage -- mortality rate scheme of analysis could be employed.

Results based on limited data at this 1000-second exposure showed that the most probable voltage limits for 99% survival were as follows: about 400 volts for the impregnated units, and about 200 volts for the unimpregnated units Temperatures as high as 150°C were shown to have little effect at this level of survival.

These ratings are higher than those which would be applied to these capacitors if they were rated by the standard long-life test.

Three characteristics of capacitors are commonly measured as a function of temperature, namely, the dissipation factor, the megohm-microfarad product, and the capacitance. In order to complete the list of characteristics required to intelligently use capacitors as fuze components in ammunition items such as guided missiles and mortars, their "short time" operating voltage should also be measured as a function of temperature. Based on considerations of reliability and operating life of a capacitor as a fuze component, the time-period of study should be limited to a maximum of 1000 seconds.

The object of the experiments reported herein was two-fold: (1) to determine a method by which capacitors could be rated for voltage-vs-temperature characteristics for a short period of time, and (2) to determine if there was a significant difference between the rating so obtained and the long-time voltage-vs-temperature rating now commonly used by most capacitor vendors.

The scheme of analysis which was chosen for treatment of test data is known as the dosage-mortality method and stems from problems having to do with the lethality of varying dosages of drugs. Friedman 1/ has summarized the method of analyzing such data. It involves the assumptions that: (1) each capacitor is characterized by a certain voltage at which it will fail and (2) that the voltages required for failure, or some function of those voltages, are normally distributed among the capacitors. Observed proportions of capacitors failing at a certain voltage can be converted into an estimate, y, of Y by means of tables of the normal probability integral. Y is defined as follows:

$$F(\log V) = 1/\sqrt{\pi 2}\,)\int_{Y}^{\infty} \exp(-\frac{x^2}{2})\,dx$$

where:

$Y = (\log V - \mu)/\sigma$

V = applied voltage

μ = true arithmetic mean of log V

σ = true standard deviation of log V

Y is linearly related to log V by:

$$Y = -\frac{\mu}{\sigma} + \frac{1}{\sigma} \log V$$

The method of calculating the "best estimate" of μ and σ, and consequently of determining Y from experimental data, is given by Bliss 2/. The method of calculating the confidence limits is also given by Bliss.

EXPERIMENTAL METHODS. Two types of commercial metal-cased capacitors were used in this work; both had a 0.25-mil-thick single layer of polyethylene terephalthate as the dielectric but one was impregnated and the other was not. The impregnant was polyisobutylene containing a small percentage of additive. The nominal value of the capacitors was 0.05 μf.

A block diagram of the electrical circuit used to perform the tests described below is shown in Figure 1.* The high-voltage supply was variable from 0 to 15,000 volts. The relay was arranged to interrupt the high voltage and stop the timer when the current reached 3 milliamperes. A peak reading voltameter was used in order to allow sufficient time to read the maximum voltage reached. The meter had a response better than 100 volts per millisecond.

A sensitivity test was first performed on both the impregnated and the unimpregnated capacitors in order to establish the voltage ranges over which voltage-dosage-mortality tests could be carried out. A lot of each type of capacitor was divided into four groups of ten units each. A voltage which increased at the rate of 100 volts per millisecond was then applied to each capacitor in the group until failure occurred. Failure was defined as the voltage point at which a current of three milliamperes flowed. One group of each type of capacitor was tested at each of the following four temperatures: 23°, 85°, 125° and 150°C. For each group, the average breakdown voltage and the standard deviation were calculated. The left-hand portions of Tables 1 and 2 show the resultant test data for the impregnated and unimpregnated capacitors, respectively.

Voltage-failure tests were then performed on both the impregnated and unimpregnated capacitors in the following manner. A lot of each type of capacitor was divided into several groups of ten units each. Groups were tested at each of the following temperatures: 23°, 85°, 125°, 150° C.

---

* Figures have been placed at the end of the article.

Each capacitor was subjected to a given voltage for 1000 seconds. The actual voltages applied to each group at a given temperature were varied in order to cover the range from no failures to about 100 percent failure. In each instance the number of failures per group was recorded. Those units which survived this test were then restressed at the same temperature but higher voltage. The right-hand portions of Tables 1* and 2 show the resultant test data for the impregnated and unimpregnated capacitors, respectively.

The proportions of capacitors that failed at the various voltages for a given temperature were then plotted in cumulative frequency form and a best-fitting straight line was calculated. Zero and 100% values which were used for calculating the straight line cannot be shown on the normal proabbility scale. Figures 2 through 5 show the fitted functions, as well as the 90% confidence limits, for the impregnated capacitors at the four test temperatures, while Figures 6 through 9 show similar data for the unimpregnated capacitors. Data from these curves were then used to plot the voltage-temperature curves at 1%, 10% and 50% failure which are shown in Figures 10 and 11 for the impregnated and unimpregnated capacitors, respectively.

The length of time required for each capacitor to fail at a particular condition was also recorded but these data are not included herein for reasons given in the following section.

DISCUSSION. The statistical method used herein does not yield as precise an estimate of the extremes of the distribution so well as it does the estimate of the central value. This is shown in Figures 2 through 9 by the spread of the confidence limits at the extremes.

The voltages required for failure of both types of capacitors investigated herein varied widely as shown by the slope of the best-estimate line of Figures 2 to 9, and also by the large standard deviation, S, in the sensitivity tests of Tables 1 and 2. Because of the large variations in the former tests, any effects due to time were largely masked and no positive statement can be made concerning such effects.

The sensitivity test for which values are reported in Tables I and II is similar time-wise to the 15-second flash test 3/ used for acceptance testing by industry. This sensitivity test yielded average breakdown voltages ($\overline{V}$) which were enough larger than the 50% points of Figures 10 and 11 to be considered significantly different. Therefore, if a sensitivity test of the type were used to determine the rated voltage for 1000-second operating life it would yield too high a value, at least at low temperatures. At increasing temperatures, the two values tended to converge.

As shown in Figures 10 and 11, the most probable estimates of the voltage limits for 99% survival were as follows: about 400 volts for the impregnated units and about 200 volts for the unimpregnated units. These figures also show that, for 99% survival, the voltage limits were not affected by temperatures increasing to as high as 150°C but that, for 50%

* Tables can be found at the end of this article.

survival, the voltage limits were apparently decreased by increasing temperatures.

The maximum average breakdown voltage (1953 volts) obtained at room temperature in these tests gave a dielectric strength of 3.1×10^6 volts per cm which is only approximately ½ of the dielectric strength measured at Massachusetts Institute of Technology 4/ for polyethylene terephthalate itself. Hence, it is concluded that, if the number of imperfections in these capacitors could be reduced, their voltage rating could be increased.

Long-life tests 5/ have not been made on these capacitors. Normally 0.25-mil polyethylene terephthalate capacitors are rated at 150 to 200 bolts dc for long life. However, long-life testing is time consuming, especially if done over a wide temperature range, and would probably yield a value which is too low for maximum space utilization when the operating life is 1000 seconds. On the other hand, this voltage-dosage-mortality test method does yield values for voltage-temperature ratings in a comparatively short period of time. The test time is compatible with the expected operating life.

This test will be modified in an effort to acquire meaningful voltage-time data.

ACKNOWLEDGMENT. The author wishes to thank Joseph Kaufman of the Office of the Chief of Ordnance for outlining this task, and Badrig M. Kurkjian, Margaret A. Hamil and Victor Labolle of the Diamond Ordnance Fuze Laboratories for their assistance in the statistical treatment of data.

BIBLIOGRAPHY

1/ C. Eisenhart, M. W. Hastay, and W. Allen Wallis, "Selected Techniques of Statistical Analysis for Scientific and Industrial Research and Production and Management Engineering", 1st Ed., Chapter 11 by Milton Friedman, page 342, McGraw-Hill Book Co., Inc., New York (1947).

2/ C. I. Bliss, "The Calculation of the Dosage-Mortality Curve", Ann. Appl. Biol., 22, 134-137 (1935).

3/ Military Specificiation MIL-C-25A, "Capacitors, Fixed, Paper-Dielectric, Direct-Current (Hermetically Sealed in Metallic Cases)", 9 March 1953, paragraph 4.6.2.

4/ Y. Inuishi and D. A. Powers, "Electric Breakdown and Conduction through Mylar Films", Technical Report 112, Laboratory for Insulation Research, Massachusetts Institute of Technology, December, 1956.

5/ Reference (3), paragraph 4.6.13.1.

Table 1. -Failure a/ of ½-mil single-layer polyethylene terephthalate capacitors, with impregnant, as a function of voltage and temperature

Temp. °C	Breakdown b/ voltage, volts: $\overline{V}$	s c/	s/$\overline{V}$	Failure -- Voltage-dosage-mortality test (see key d/ below)	Voltage, volts: 200	400	500	600	650	800	1100	1400
23	1953	311.9	0.16	(1)	--	--	10	--	10	10	10	10
				(2)	--	--	0	--	0	40	60	60
				(3)	--	--	800	--	950	1100	1400	1700
				(4)	--	--	50	--	40	33	50	75
85	1899	421.7	0.22	(1)	--	--	10	--	10	7	10	--
				(2)	--	--	20	--	20	86	80	--
				(3)	--	--	800	--	950	1100	1400	--
				(4)	--	--	50	--	50	0	100	--
125	819	475.6	0.58	(1)	10	--	10	--	10	10	--	--
				(2)	0	--	20	--	60	100	--	--
				(3)	500	--	800	--	750	--	--	--
				(4)	40	--	100	--	75	--	--	--
150	618	348.8	0.564	(1)	10	10	10	10	--	--	--	--
				(2)	0	30	50	90	--	--	--	--
				(3)	400	700	800	900	--	--	--	--
				(4)	40	100	100	100	--	--	--	--

a/ Failure was defined by a current of 3 milliamperes.
b/ Voltage was increased at the rate of approximately 100 volts per millisecond until failure occurred.
c/ Standard deviation
d/ Key: (1) top, (3) left, (4) right, (2) bottom of each cell

(1) Number tested
(2) Percent failed in (1)
(4) Percent failing in retests

Table 11. -Failure a/ of ¼-mil single-layer polyethylene terephthalate capacitors, without impregnant, as a function of voltage and temperature

Temp. °C	Breakdown b/ voltage, volts: $\overline{V}$	S c/	$S/\overline{V}$	Key d/	Failure --- Voltage-dosage-mortality test (see Key d/ below) Voltage, volts: 150	200	300	450	600	900	1200
23	864	225	0.26	(1)	--	10	10	--	10	10	10
				(2)	--	0	30	--	50	60	100
				(3)	--	500	600	--	900	1200	--
				(4)	--	20	43	--	40	100	--
85	958	361	0.376	(1)	--	--	10	--	10	10	--
				(2)	--	--	10	--	30	90	--
				(3)	--	--	600	--	900	1200	--
				(4)	--	--	22	--	43	100	--
125	889	407	0.458	(1)	10	--	10	--	10	10	10
				(2)	0	--	10	--	90	100	100
				(3)	450	--	600	--	900	--	--
				(4)	20	--	78	--	100	--	--
150	664	237	0.357	(1)	10	--	10	10	10	10	--
				(2)	10	--	30	40	100	100	--
				(3)	450	--	600	750	--	--	--
				(4)	89	--	71	83	--	--	--

a/ Failure was defined by a current of 3 milliamperes.
b/ Voltage was increased at the rate of approximately 100 volts per millisecond until failure occurred.
c/ Standard deviation
d/ Key:

	(1)	
(3)		(4)
	(2)	

(1) Number tested
(2) Percent failed in (1)
(3) Voltage at which the units surviving number (1) were retested, volts
(4) Percent failing in retests

TITLES OF FIGURES

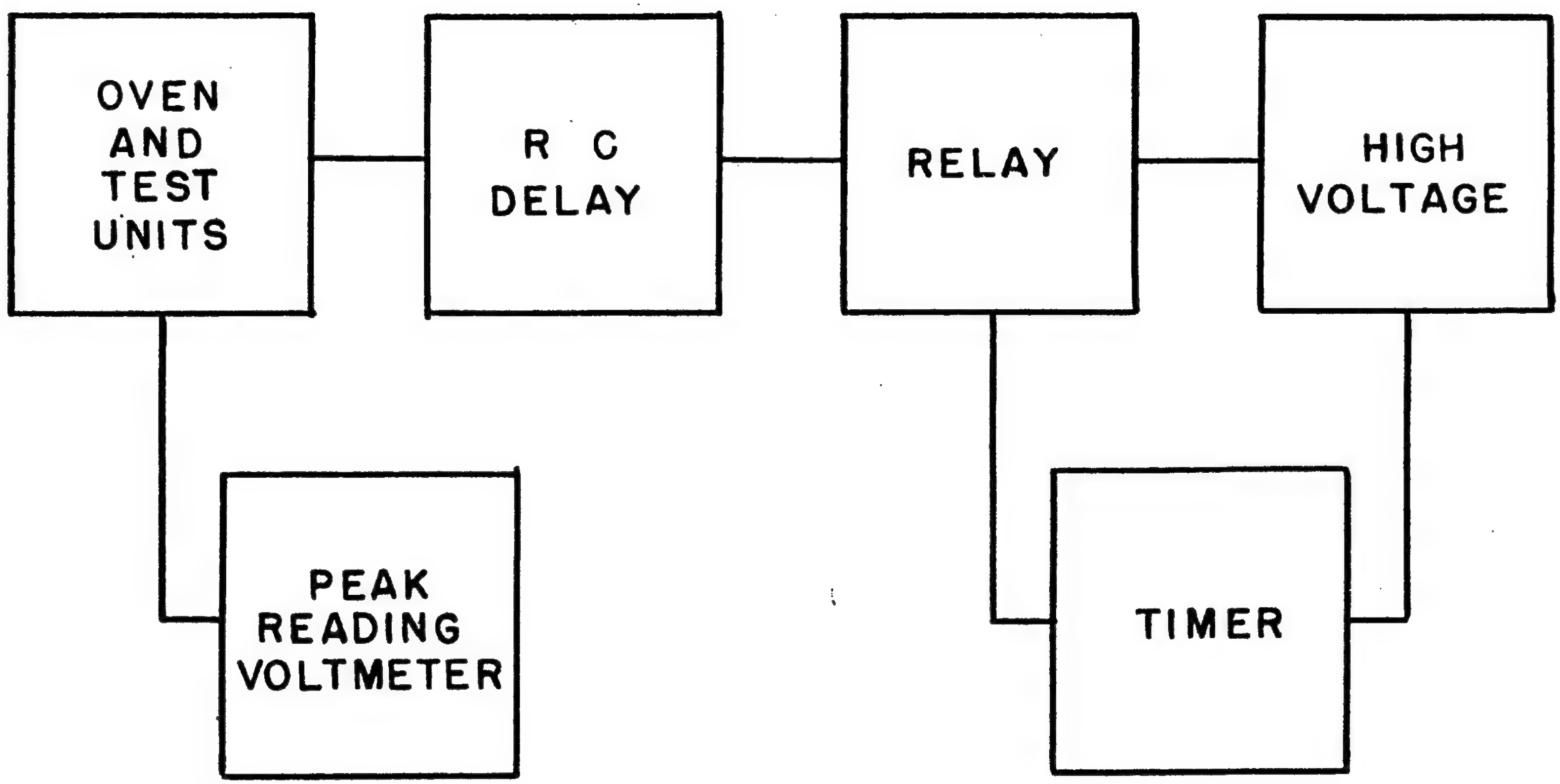


Figure 1. Block diagram of apparatus for voltage-dosage-mortality tests.

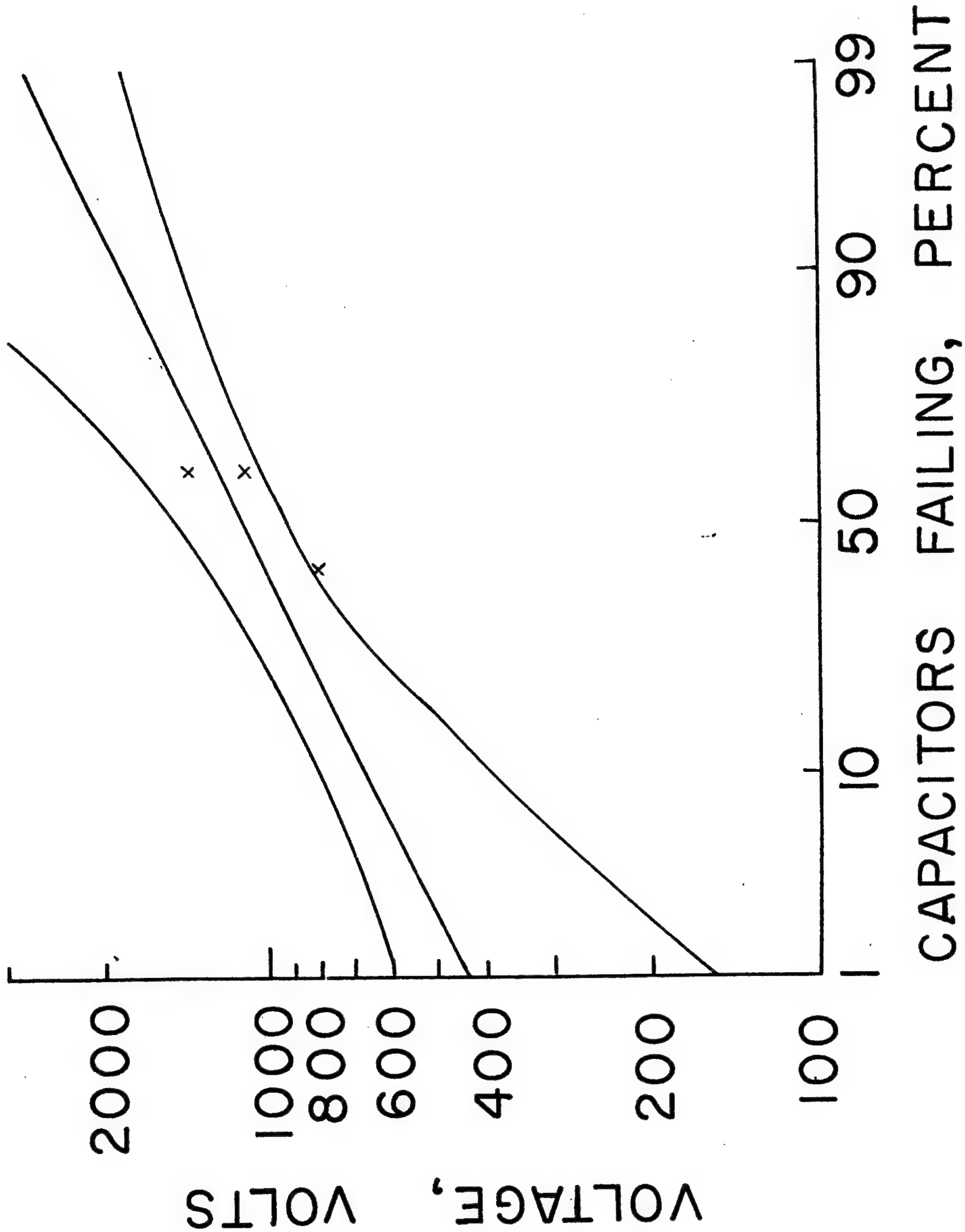


Figure 2.-Cumulative frequency of failure vs. log voltage, and 90-percent confidence levels, at 23°C (73°F) for impregnated capacitors using 0.25-mil-thick polyethylene terephthalate

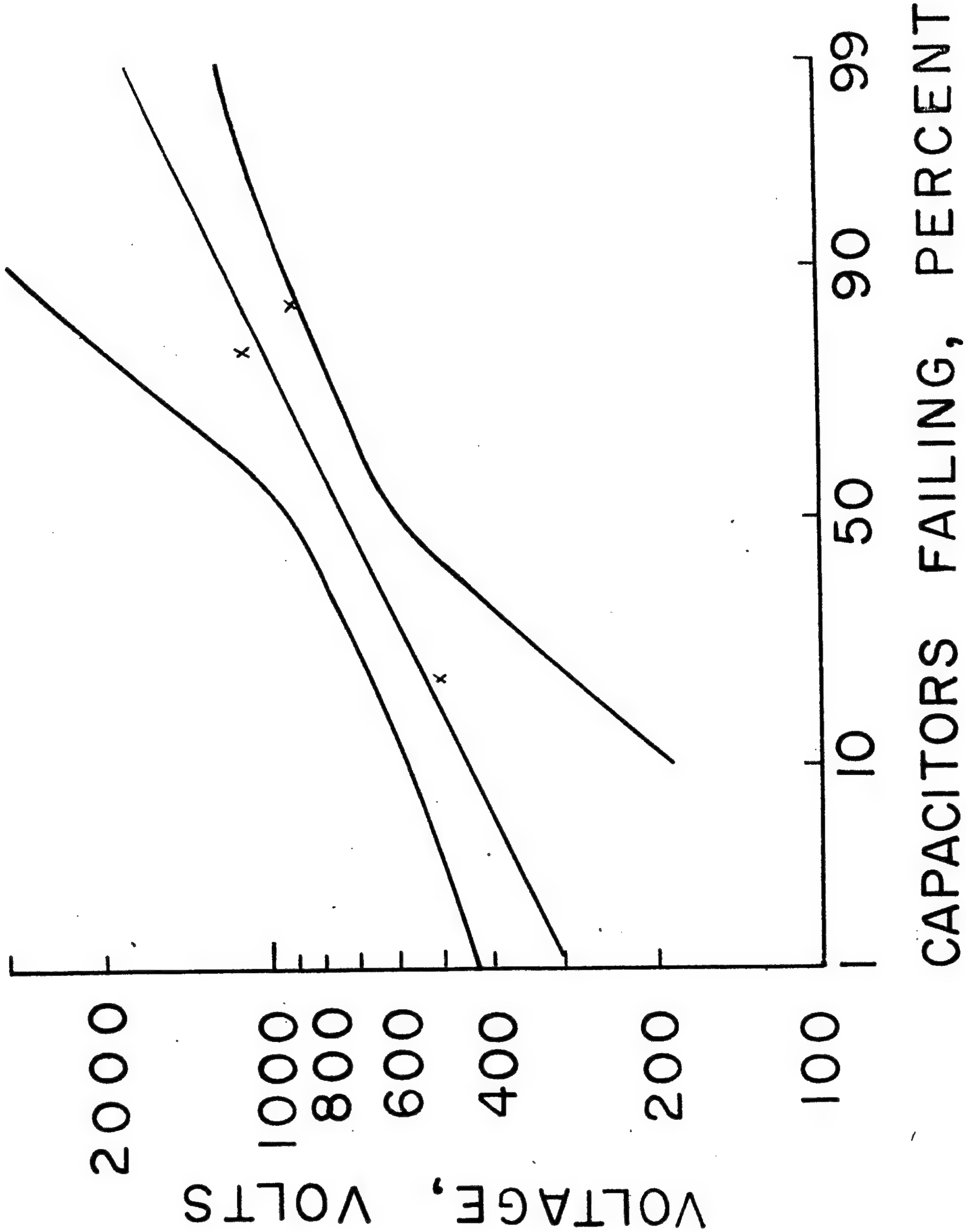


Figure 3. Cumulative frequency of failure vs. log voltage, and 90-percent confidence levels, at 85°C (185°F) for impregnated capacitors using 0.25-mil-thick polyethylene terephthalate

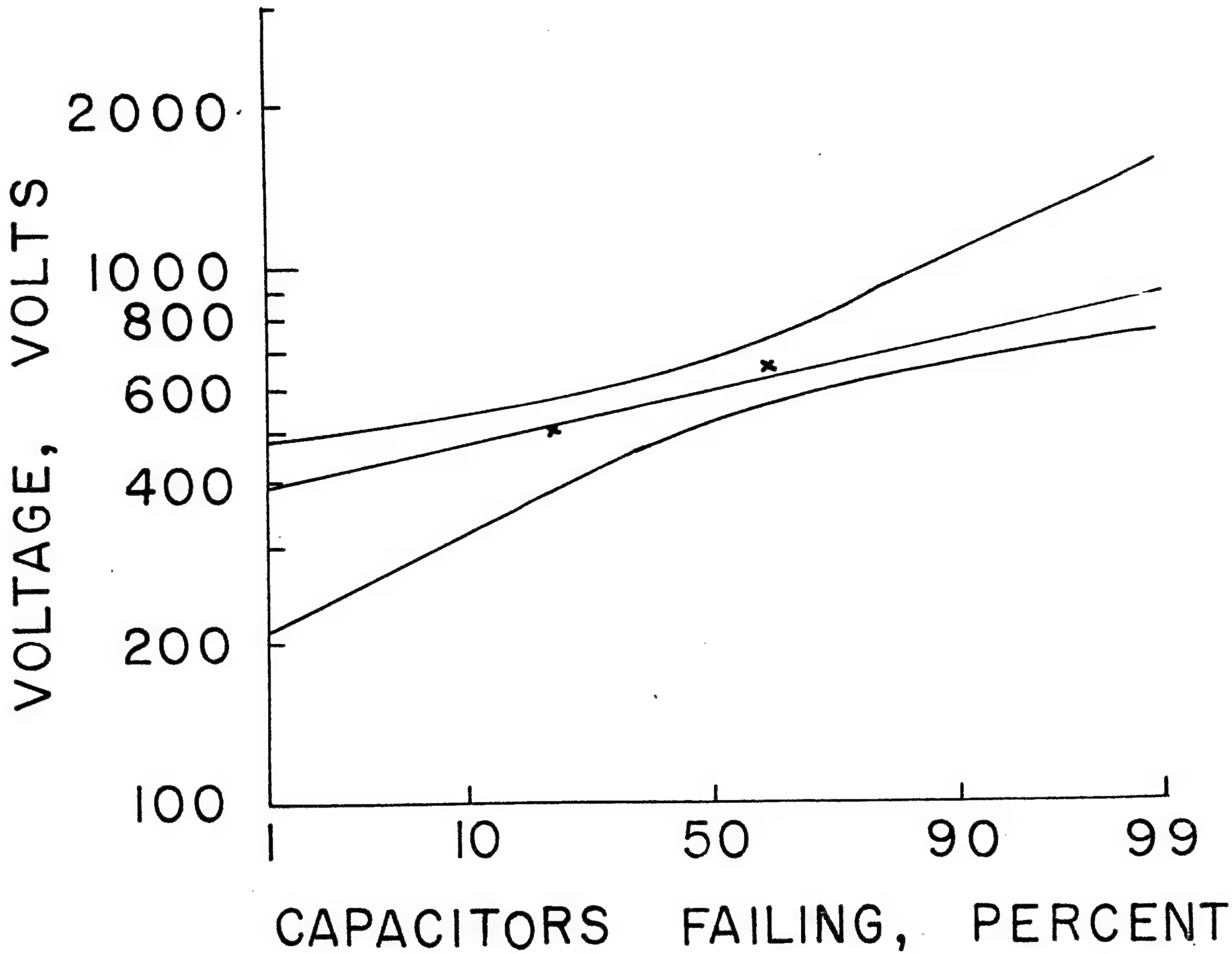


Figure 4. Cumulative frequency of failure vs. log voltage, and 90-percent confidence levels, at 125°C (257°F) for impregnated capacitors using 0.25-mil-thick polyethylene terephthalate

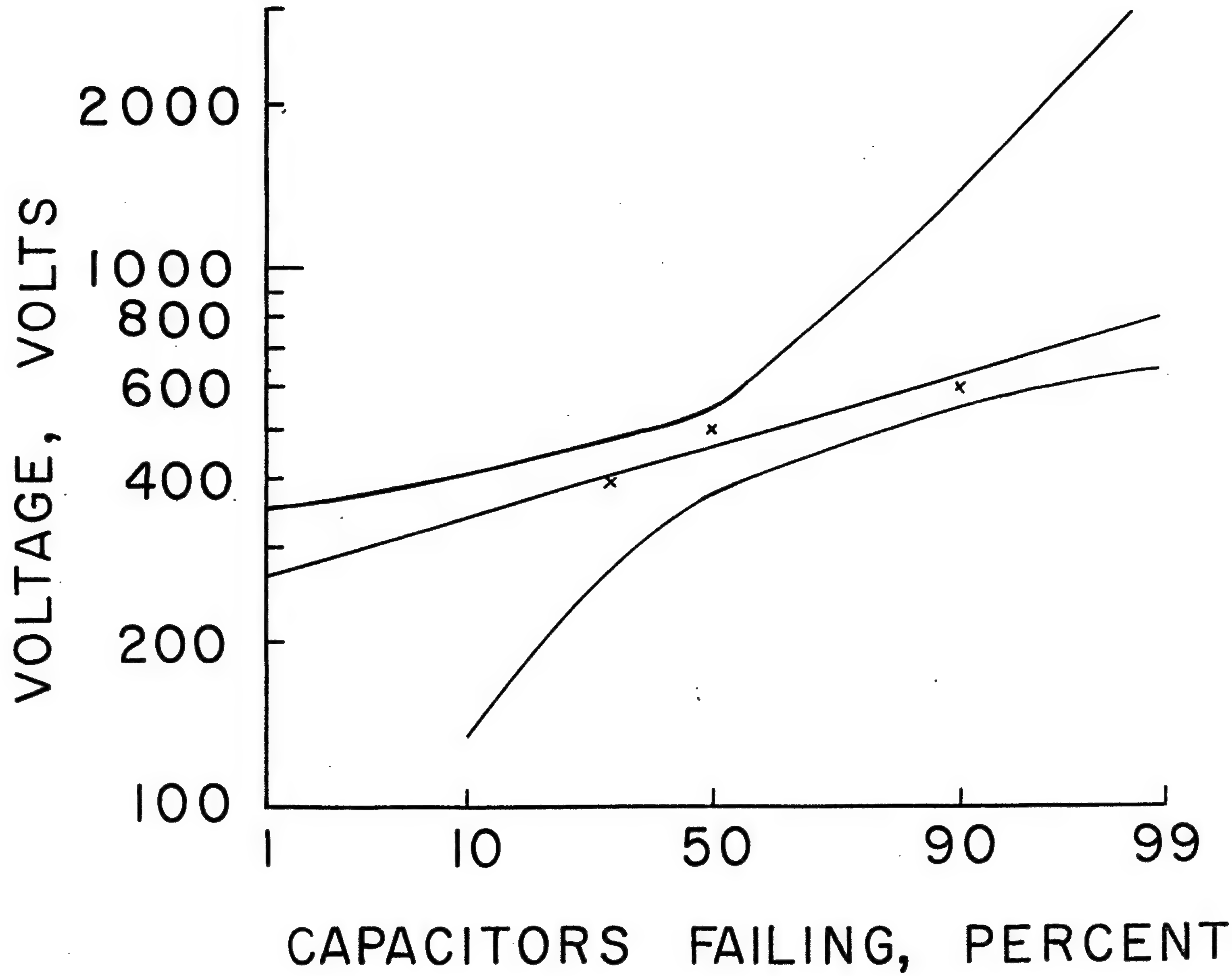


Figure 5. Cumulative frequency of failure vs. log voltage, and 90-percent confidence levels, at 150°C (302°F) for impregnated capacitors using 0.25-mil-thick polyethylene terephthalate

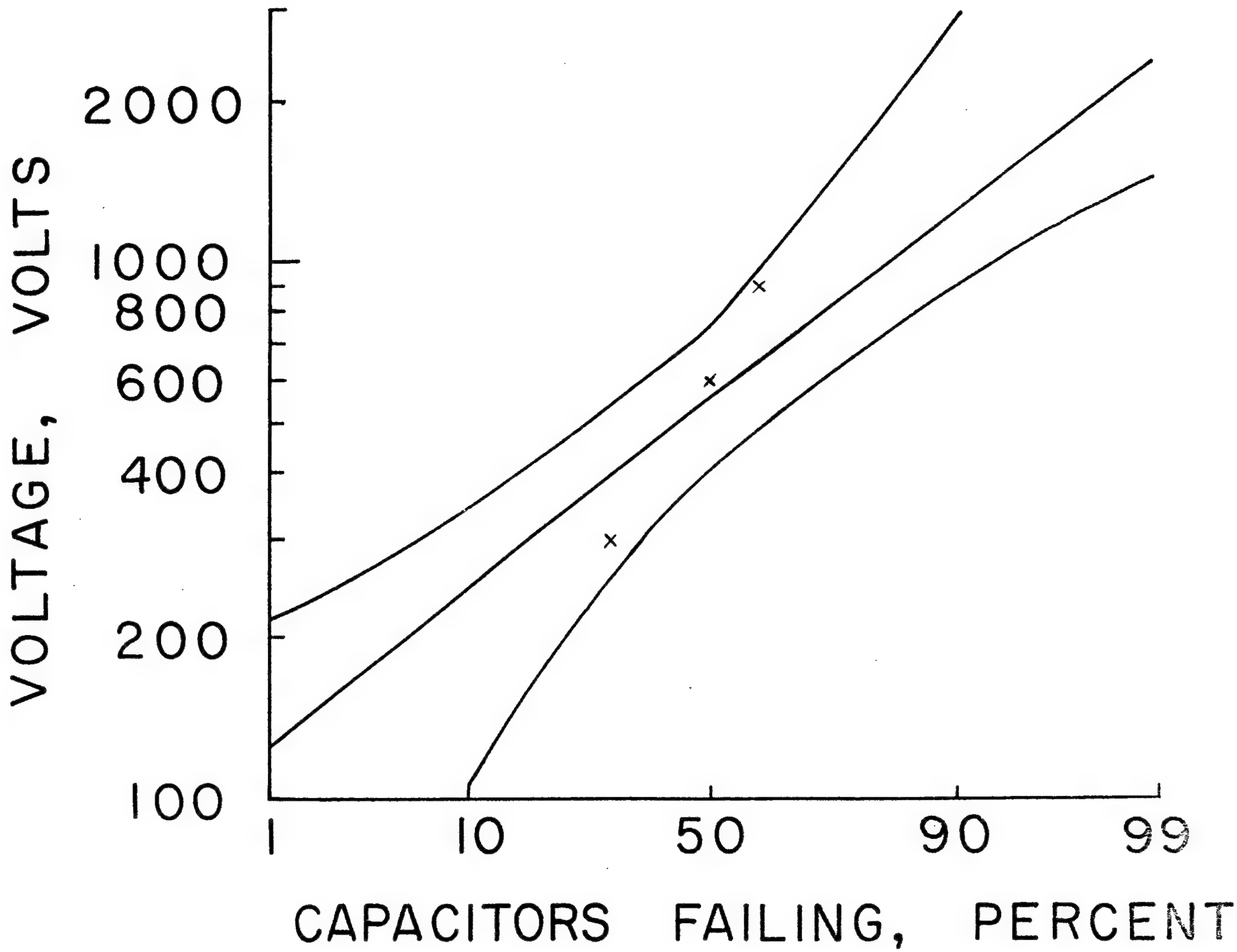


Figure 6, Cumulative frequency of failure vs. log voltage, and 90-percent confidence levels, at 23°C (73°F) for unimpregnated capacitors using 0.25-mil-thick polyethylene terephthalate

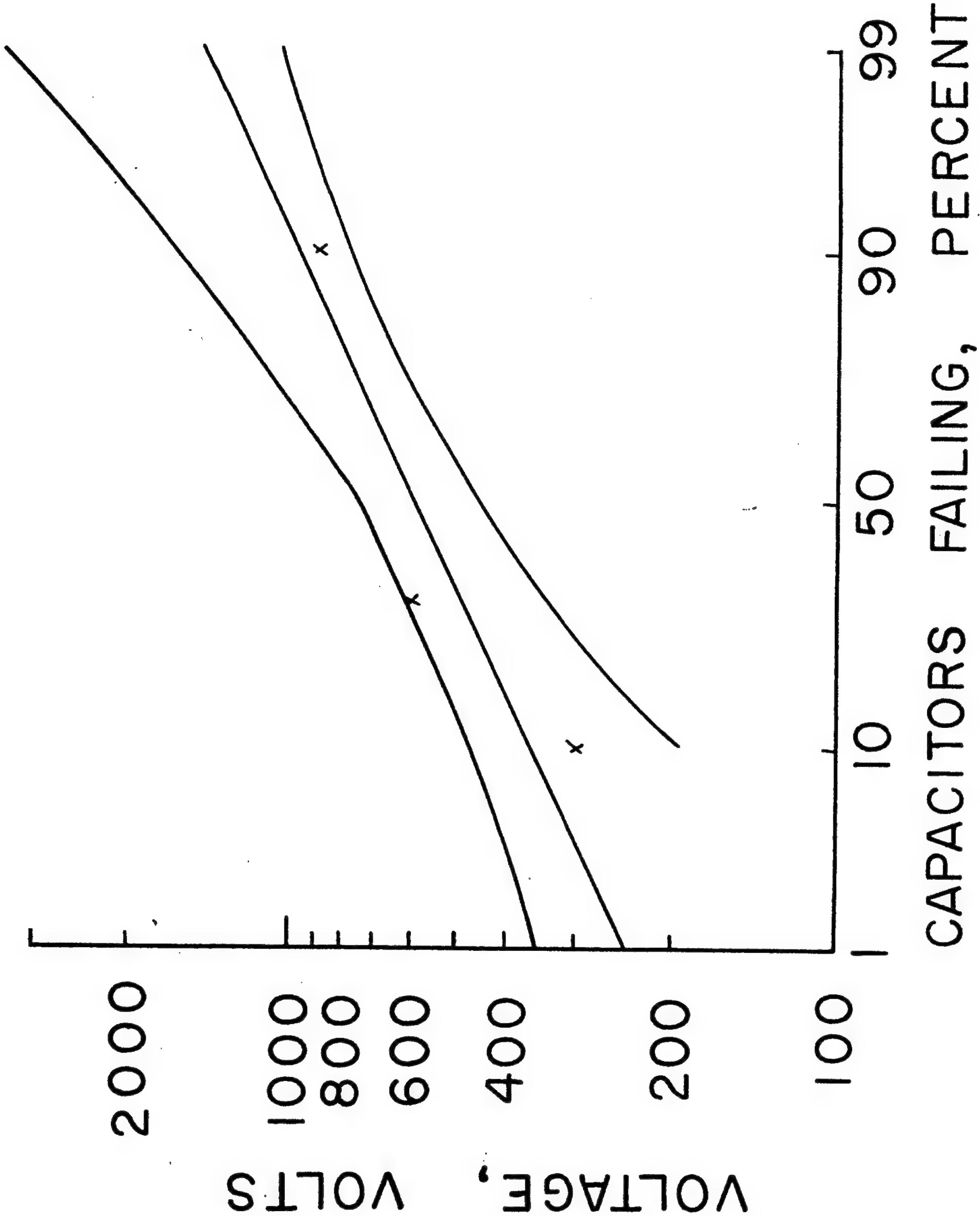


Figure 7. Cumulative frequency of failure vs. log voltage, and 90-percent confidence levels, at 85°C (185°F) for unimpregnated capacitors using 0.25-mil-thick polyethylene terephthalate

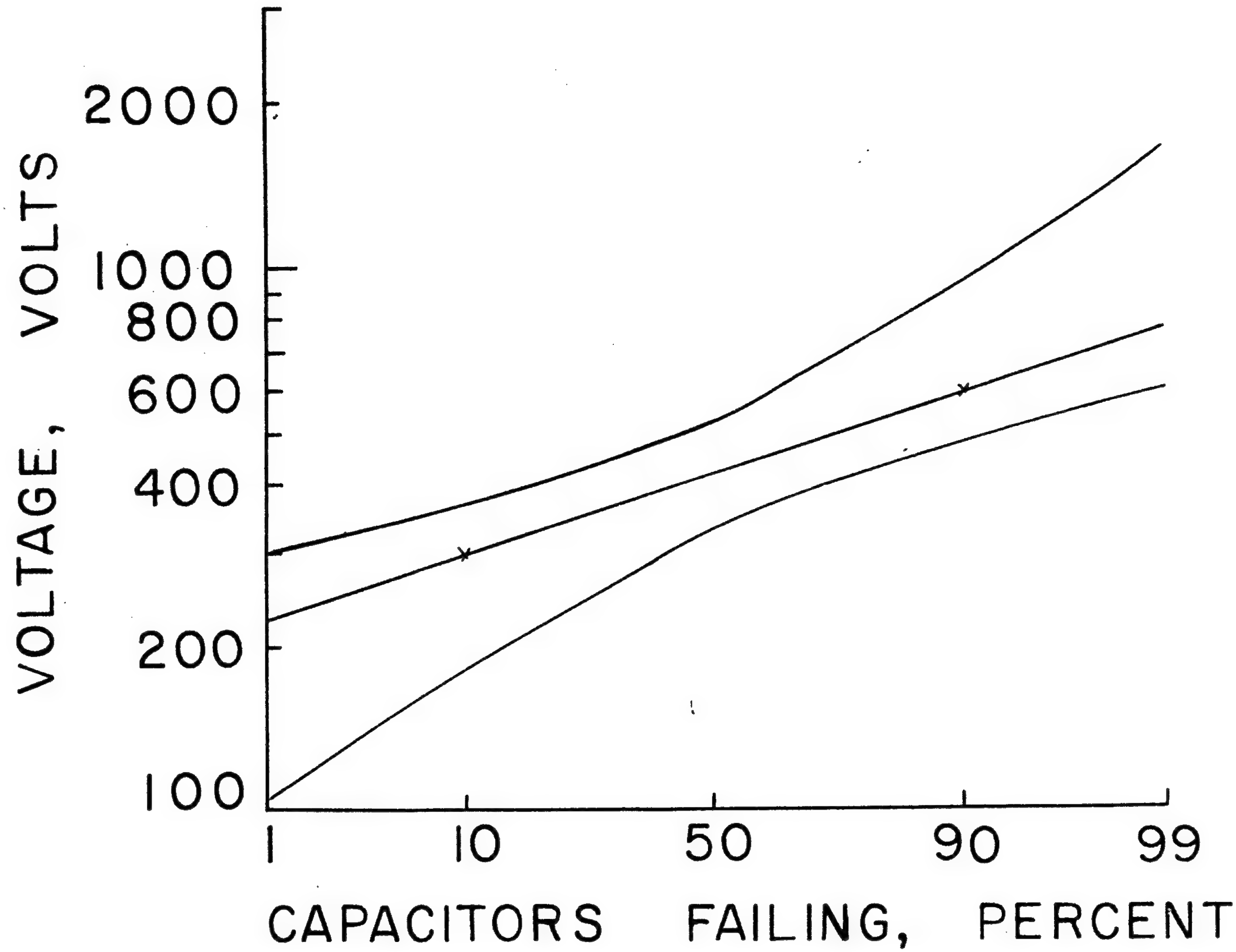


Figure 8. Cumulative frequency of failure vs. log voltage, and 90-percent confidence levels, at 125°C (257°F) for unimpregnated capacitors using 0.25-mil-thick polyethylene terephthalate

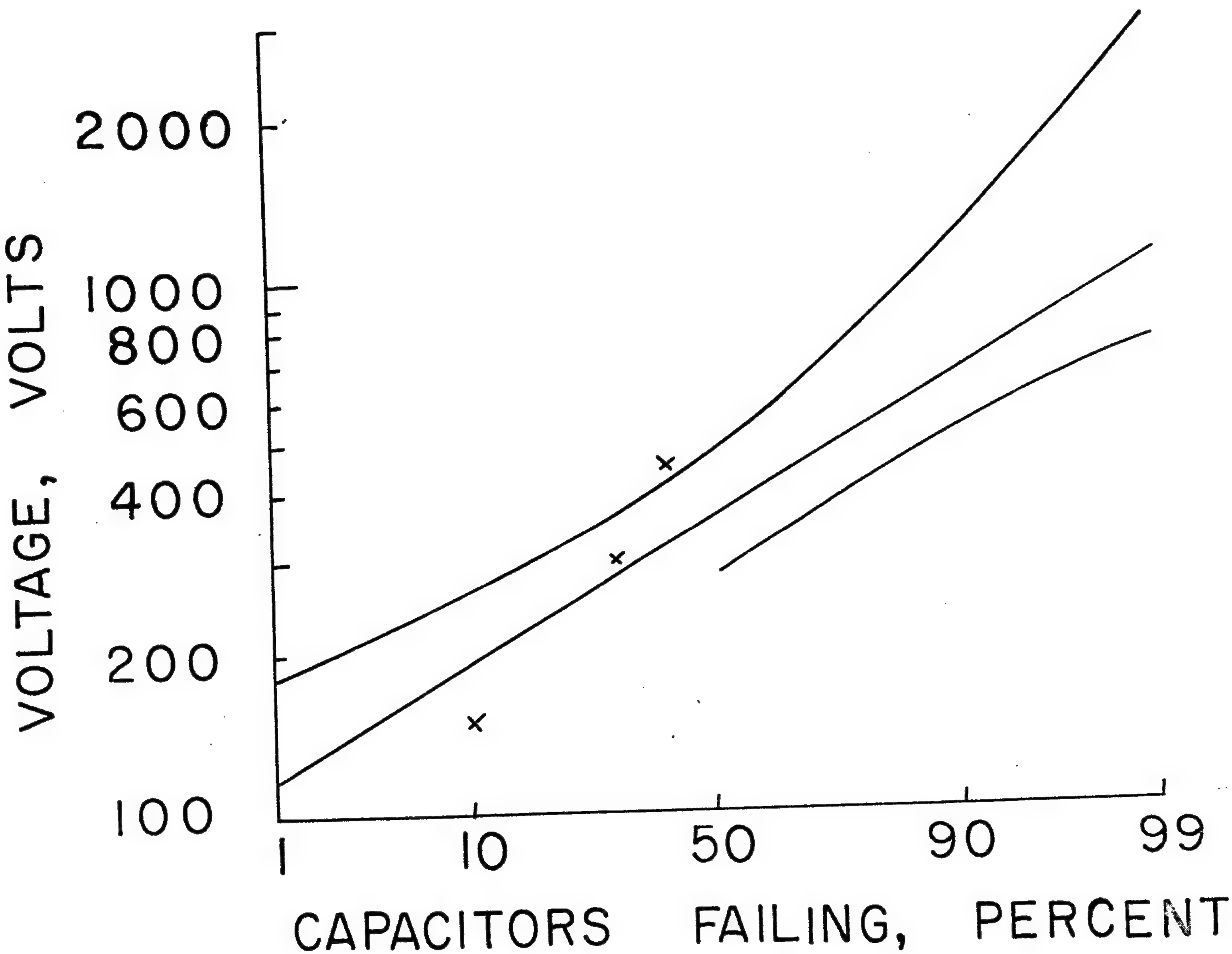


Figure 9. Cumulative frequency of failure vs. log voltage, and 90-percent confidence levels, at 150°C (302°F) for unimpregnated capacitors using p.25-mil-thick polyethylene terephthalate

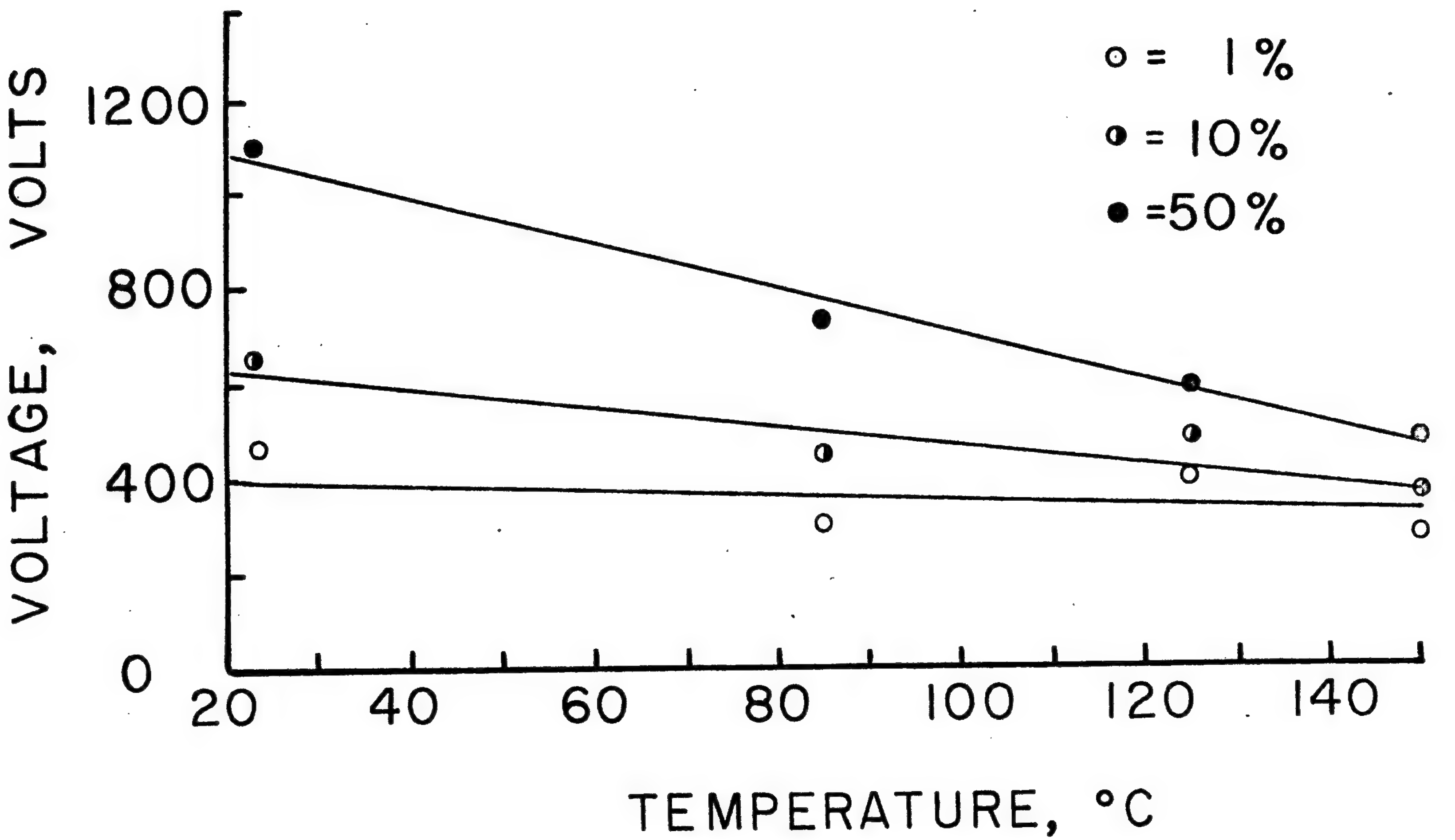


Figure 10. Voltage required for 1, 10, and 50-percent failures vs. temperature for a lot of impregnated capacitors using 0.25-mil-thick polyethylene terephthalate as the dielectric.

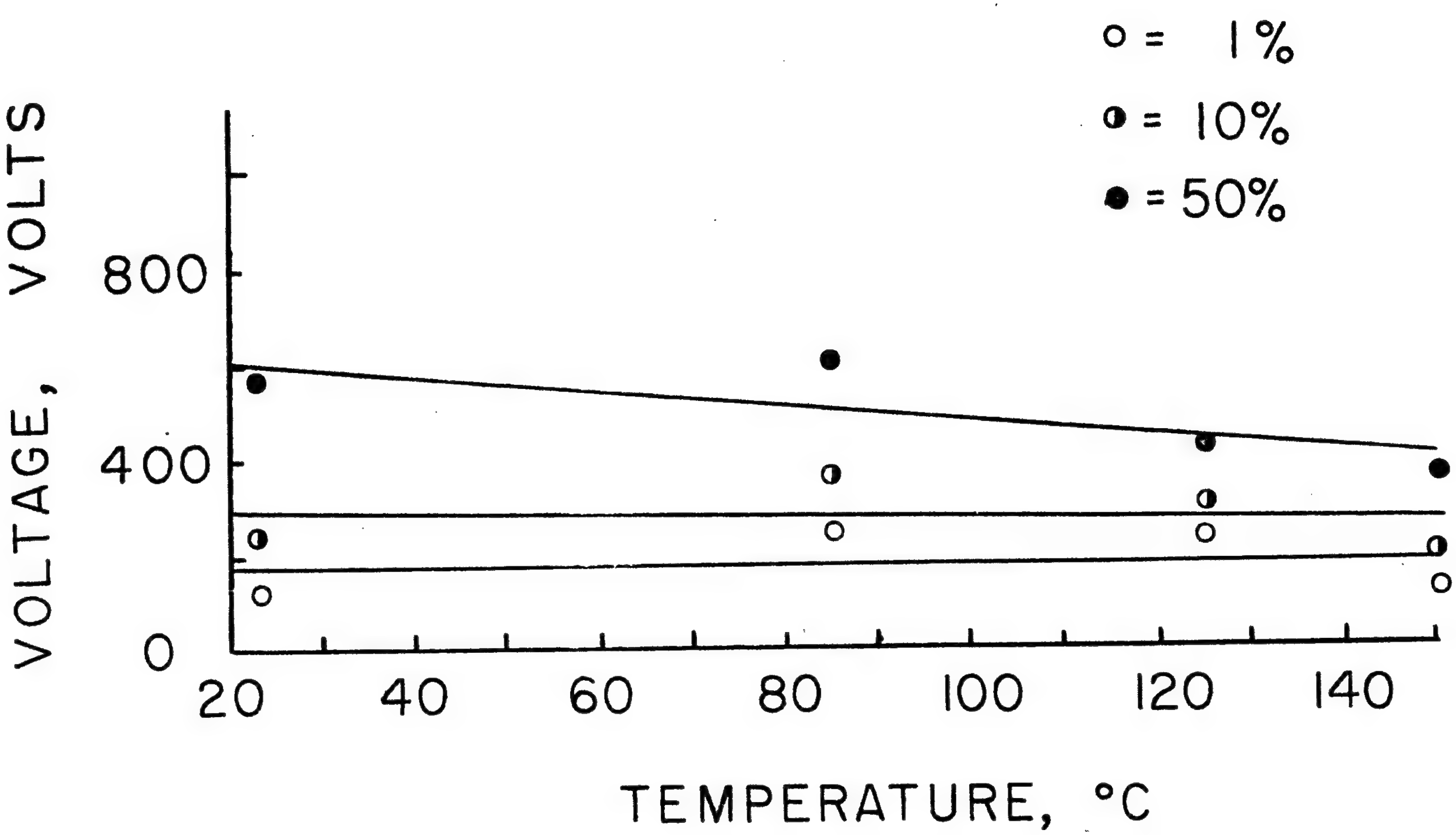


Figure 11. Voltage required for 1, 10, and 50-percent failures vs. temperature for a lot of unimpregnated capacitors using 0.25-mil-thick polyethylene terephthalate as the dielectric.

THE DESIGN OF CONTROLLED SIMULATION EXPERIMENTS

Melvin D. Springer
Combat Operations Research Group

I INTRODUCTION. Problems attacked by simulation procedures are generally of a highly complex nature with stochastic features that enter in various ways. Consequently, care, must be taken to obtain results which are sufficiently accurate to be useful. This exercise of caution must begin with the design of the experiment. I should like to discuss a few of the central factors which enter into the design of a controlled simulation experiment.

II THE SIMULATION MODEL. The first factor which we encounter in this type of problem is, of course, a simulation model. It goes without saying that the simulation model must be adequate. That is, the simulation model - no matter what its structure - must produce estimates which are consistent with the results produced from its physical counterpart. The structure of the model depends upon various things, such as the type of sampling employed (synthetic sampling or Monte Carlo), the type of process involved (e.g., Markovian or non-Markovian), etc.

Consider first the effect of the type of sampling employed. If straightforward synthetic (experimental) sampling is used, then it is important that our model adhere strictly to the physical situation which it approximates. Thus, in a simulated tank battle designed to compare the vulnerability of two types of antitank weapons under certain terrain conditions, the tanks may at times be firing against visible targets and sometimes against invisible targets whose general position is indicated by smoke of flash. With synthetic sampling,it is imperative that we incorporate into our model reasonbly accurate probability distributions for kills achieved by tanks against both visible and invisible targets. On the other hand, it is possible to introduce certain distortions into the simulation model which cause it to deviate from its physical counterpart without invalidating the results. For instance, in the above example, if most of the tank firing is against invisible targets which have a low probability of being killed, we might wish to increase (distort) the probability of a tank's killing an invisible target in order to get, without an undue amount of sampling, a sufficient number of kills to permit satisfactory analysis. In order to do this and yet avoid bias, the results must be properly weighted. The term properly weighted here covers a multitude of sins, but in the simplest situations (not necessarily here) the weights used are inversely proportional to the factor by which the probability of occurrence of the event was distorted. Any procedure in which the sampling has been modified in some such fashion will be referred to as a Monte Carlo procedure as distinguished from a straightforward synthetic sampling procedure devoid of all distortions. The main justification for resorting to Monte Carlo techniques is that they will, if used judiciously, yield valid, unbiased results with considerable reduction in variance, hence requiring less sampling to attain a specified degree of precision. Some simulation problems do not lend themselves to such Monte Carlo procedures; and for those which do, a good bit ingenuity, thought, and invesitagion of the problem are required in order to select a Monte Carlo procedure which will bring about a substantial reduction in variance. We shall later give some examples indicating how Monte Carlo procedures achieve variance reduction in some very simple cases.

If the problem being analyzed by simulation methods is of a Markovian nature, a Markovian simulation model may be used with consequent simplification of the resultant analysis. A Markov process is one in which the future development is influenced only by the present state and is independent of the way in which the present state has developed. The processes of classical mechanics are of this type, as contrasted with processes in the theory of plasticity, where the whole past history of the system influences its future. In stochastic processes, the future is never uniquely determined, but we have probability relations enabling us to make predictions. For Markov chains, the probability relations relating to the future depend on the present state, but not on the manner in which the present state has emerged from the past. For simulation problems for which a Markovian model is applicable, certain advantages accrue insofar as the analysis of the problem is concerned, particularly if we are dealing with a problem whose magnitude requires the use of a high-speed computer. That is, once the appropriate Markovian model has been set up, there are straightforward procedures for determining the probability that the system passes from state j to state k in exactly n steps. These procedures are equivalent to finding the eigne values of a matrix, for which routine procedures exist in almost all computer installations.

III <u>ATTAINING STABILITY</u>. In order to obtain results sufficiently stable to be useful, i.e., estimates with reasonably small variances, in the types of problems dealt with by simulation, it is often necessary either to employ variance-reducing sampling techniques or to work with very large samples. Let us consider each of these approaches toward attaining stability.

A. Variance-Reducing Techniques

In order to illustrate the general nature of these techniques, we shall use them to solve a very simple problem. As John Tukey has so aptly remarked [1], the only good Monte Carlos are dead Monte Carlos-the ones we don't have to do. Nevertheless, while the problem we are about to cite is almost trivial, its solution by the application of Monte Carlo techniques does serve to illustrate the principles behind these techniques.

The example we shall consider is the problem of calculating the probability of obtaining a total of three when two ordinary dice are tossed. Since each die is a standard one with six faces labeled from one to six, each face has the same probability (1/6) of being on top. The problem has, obviously, a simple analytical solution. Any particular combination of the dice has a probability of occurrence equal to $(1/6)^2$. Since there are two combinations of faces resulting in a total of three (one-two and two-one), the probability of getting a three in a random toss is $2 \times (1/6)^2$ of 1/18.

In attacking this problem by straightforward synthetic sampling, one would simply toss the dice N times, count the number (n) of successes (threes) and then estimate the probability (p) of success by

$$\hat{p} = \frac{n}{N} \tag{1}$$

Clearly, the estimate obtained in this way is subject to random sampling fluctuations giving rise to a statistical error usually measured by the standard deviation σ . In this problem,

$$\sigma = \sqrt{\frac{p\,(1-p)}{N}}\,, \tag{2}$$

or expressed percentagewise

$$\frac{100\,\sigma}{p} = 100\sqrt{\frac{1-p}{Np}} \tag{3}$$

One way of reducing this error is to increase N. There are other ways in which this error can be reduced, namely, variance-reducing techniques. Some of these techniques described by Herman Kahn (and throughout this section we shall lean heavily on Kahn's article [2]) are: importance sampling, Russian roulette and splitting, use of expected values (combination of analytic and probabilistic methods), correlation and regression, systematic sampling, and stratified sampling. Kahn states that the first three of these seem to have found particular and specialized usefulness in Monte Carlo applications. Following Kahn, we shall illustrate the general nature of these three techniques by applying them to the solution of the simple problem posed and solved analytically above.

1. Importance Sampling

If we can somehow increase the effective value of p, it is clear from (3) that the percentage error will be reduced. If we bias the dice by "loading" them so that the probability that a one or two comes up is twice as great as usual, then the probability of getting a three is increased error is then reduced by approximately a factor of two. Clearly, (1) can no longer be used to estimate p, but most be replaced by

$$\hat{p} = \frac{1}{4}\,\frac{n}{N} \tag{4}$$

That is, we must apply a weighting factor of 1/4 to our original estimate of p to remove the distortion introduced by the biased sampling.

This illustrates the general idea of importance sampling, which consists of drawing samples from a distribution other than the one suggested by the problem and then to carry along an appropriate weighting factor, which when multiplied into the final results, corrects for having used the wrong distribution. The improvement results from the fact that the biasing is done in such a way that the probability of the sample's being drawn from an "interesting" region is increased while the probability of its being drawn from an "uninteresting" region is correspondingly decreased. It is perfectly legitimate to carry the bias to the limit, i.e., to increase the probability of getting a one or a two by a factor of three, making the probability of obtaining one of these numbers 1/2 and the probability of obtaining any other number zero.

We can, however, do even better than that. We can, for example, toss the dice one at a time and bias them differently, letting the biasing of the second die depend on the outcome of the first throw. This could be done as follows:

a. Increase the probability of getting a one or a two on the first die by a factor of three, thus decreasing to zero the probability of getting any other numbers.

b. If the first die comes up one, increase the probability of the second die coming up two by a factor of six; whereas if the first die comes up two, increase the probability of the second die coming up one by a factor of six.

If this procedure is followed every toss of the dice will yield a three; the weighting factor will then be 1/3 x 1/6 = 1/18 so that

$$\hat{p} = \frac{n}{18N} = \frac{1}{18}$$

since the number of successes (n) is equal to the number of trials (N). Inasmuch as p is now exactly equal to p, we have devised a sampling procedure which has zero variance. In principle - but not in practice - this is always possible.

In the above example, we could readily devise a sampling scheme with zero variance because we knew the answer in advance. In more complicated problems, even more than just the answer must be known before we can design a sampling scheme with zero variance. Under such circumstances, zero variance is not an amazing result. To quote Kahn [2]:

"The significance of the existence of zero variance lies not in the possibility of actually constructing them in practice but in that they demonstrate there are no 'Conservation of Cost' laws. That is, if the designer is clever, wise, or lucky he may, in choosing from the infinite number of sampling schemes available, be able to choose a very efficient one. This is in some contrast to the situation in ordinary numerical analysis. It is usually true there that once a fairly good method of doing a problem has been found, that further work or additional transformations do not reduce the cost very much, if at all. In Monte Carlo problems, however, we are assured that there is always a better way until we reach perfection."

2. Russian Roulette and Splitting

With this technique (devised and named by J. von Neumann and S. Ulam), the dice are tossed one at a time and the total number of necessary tosses reduced. Clearly, if the first die comes up three or greater, it will be impossible to get a total of three, no matter how the second die comes up. In such cases there is no point in tossing the second die; we can simply record a zero for the experiment. This makes it unnecessary to toss the second die 2/3 of the time, reducing the necessary number of tosses, on the average, by a factor of 1/3.

If the sampling is done in stages (as is frequently the case in practical problems) the sample may be examined at each stage and classified as being "interesting" or "uninteresting". Usually we wish to spend more than average amount of work on the "interesting" ones and less

effort on the "uninteresting" ones. One way to accomplish this result is to split the "interesting" samples into independent branches, thus getting more of them, and by killing off some percentage (100% in our example) of the "uninteresting" ones. The first process is known as splitting and the second as Russian Roulette. The "killing off" is done by a supplementary game of chance. If the supplementary game is lost the sample is killed; if it is won, the sample is counted with an extra weight to make up for the fact that other samples have been killed.

3. Use of Expected Values

If the sampling is done in two stages, then although we perhaps don't have sufficient insight to spell out all the combinations, we still might be able to recognize the fact that it is really unnecessary to toss the second die. That is, once the first die is tossed, it is a small matter to calculate the probability of obtaining a total of three. Thus, if the first die comes up one, the second die must come up two if the sum is to be three; the probability that this happens is 1/6. Likewise, if the first die comes up two, the second die must come up one if the total is to be three; again the probability of this result is 1/6. All other possibilities for the first die (three to six) have a zero probability of giving three.

If, then, we do not toss the second die but record the probabilities instead, the average of these probabilities is an unbiased estimate of p. This approach to the problem has a two-fold advantage: it reduces the number of tosses by a factor of two and at the same time decreases the variance, thereby making the reduced number of tosses more effective.

The illustrated technique is not merely academic, for in some practical problems analyzed by simulation, much of the variance is introduced by a part of the probabilistic problem which can be calculated analytically, while the probabilistic part which is hard to calculate analytically does not introduce much variability. The logical procedure is then to calculate analytically that which is easy and to Monte Carlo that which is hard.

While the foregoing techniques can be fantastically effective in realistic applications, Kahn injects a few words of caution with regard to their use. He states that while he is familiar with applications in which each of the three techniques has, by itself decreased the effective variance by factors of the order of 10^4 and 10^6, it is nonetheless true that if improperly used, e.g., if the intuition of the user is faulty and he does not use a reasonable design, these techniques (with the exception of the third) can be very unreliable and actually increase the variance. The experimenter usually tries to protect himself from trouble by estimating the error by means of the sample variance and then appealing to the Central Limit Theorem. While this is usually satisfactory, it can give trouble in some semi-pathological (but nonetheless real) cases. Walsh [3] cites an example which seems to be reasonable from an application viewpoint, in which the variance of the estimate $\hat{y}$ is infinte for all importance sampling functions h (x) of a given class but for which $\hat{y}$ is nonetheless a consistent estimate of y for all h(x) of the class. He suggests the use of the mean deviation instead of the variance in such cases.

We have sketched the general principles inherent in importance sampling, Russian roulette and splitting, and the use of expected values. A detailed treatment of these methods, as well as the techniques of correlation and regression, systematic sampling, stratified sampling, and conditional Monte Carlo may be found in [4], particularly in the articles by Tukey et al, and by Kahn.

B. Attaining Stability Through Use of Large Samples

We have just indicated how stability of results may be attained in some cases through the use of variance-reducing techniques. When high-speed computers are available, a logical, parallel approach would seem to be the use of techniques of sequential estimation which would provide for continual machine sampling until a sufficiently large sample was accumulated so that one obtains an estimate with a predetermined or pre-specified variance or confindence interval. In a recent paper [5], Moshman has given an excellent summary of a number of these sequential techniqes described in the literature, which can be used to estimate certain distribution parameters, namely: (1)the mean of a normal population having unknown variance, (2) the binomial parameter, (3) the mean Θ of a population whose variance $V(\Theta)$ is a finite function of Θ (4) the variance of a normal variate. Concerning these sequential techniques, Moshman states: "In every case it is possible by proper programing, and possibly some preliminary analysis, to have the computer evaluate the sample obtained thus far and determine whether additional sample units are required to obtain some specified precision. In some cases, the evaluation is done after each sample until; in the other cases, evaluation takes place at certain intervals." Since it is feasible to program these sequential techniques for a computer, they might well be used in certain simulation analyses involving evaluation of any one of the aforementioned parameters. We shall merely sketch each technique, referring the reader to Moshman's paper [5] or to the original articles for details.

1. Estimating the Mean μ of a Normal Population With Unknown Variance

Two methods are suggested: one due to Stein [7] and the other to Anscombe [8]. Stein's two-sample procedure consists of selecting an initial sample of size N_0, from which a variance estimates s^2 is determined. On the basis of N_0 and s^2, and with the aid of Student's t distribution, the experimenter can then calculate the total sample size N required to determine a confidence interval of length L for μ, corresponding to a confidence coefficient $1-\alpha$. The optimum choice of N_0 is discussed in a paper submitted by Moshman to the Annals of Mathematical Statistics [6].

Anscombe's method is a second-order asymptotic sequential procedure with a stopping rule based on the sequence of independent random variables $U_1, U_2, \ldots, U_{n-1}, Y_n$, where

$$Y_n = \frac{1}{n}\sum_{1}^{n} X_i \qquad (X_i \text{ represent the original observations})$$

$$U_i = \frac{1}{i\,(i+1)} \left(iX_{i+1} - \sum_{j=1}^{i} X_j\right)^2, \quad i=1, 2, \ldots, n-1.$$

For a confidence interval of length L, the stopping rules requires that N, the total sample size, be the least value of $n > 3$ for which

$$\sum_{i=1}^{n-1} U_i \leq \frac{L^2}{4(t_\alpha)^2} \, n\left(n - 2.676 - \frac{(t_\alpha)^2}{2}\right),$$

where t_α satisfies the equation

$$1 - \alpha = \frac{1}{2\pi} \int_{-t_\alpha}^{t_\alpha} \exp(-x^2/2)\, dx$$

The confidence interval for μ is then $(Y_n - \frac{L}{2}, Y_n + \frac{L}{2})$, with an associated confidence coefficient of approximately $1 - \alpha$. A formula for the expected sample size is available when said stopping rule is used.

2. Estimating the Binomial Parameter

In 1946, Girshick, Mosteller, and Savage [9] jointly developed a sequential procedure for estimating the parameter p, the fraction defective, of a binomial distribution. In this sequential procedure, the null hypothesis H_0: $p = p_0$ is tested against the alternative hypothesis H_1 : $p = p_1$, where $p_1 > p_0$ and denotes the fraction defective. Sampling stops when the number of defectives y is less than y_1 or when y is greater than y_2, where

$$\begin{aligned} y_1 &= -h_1 + sn \\ y_2 &= h_2 + sn \end{aligned} \qquad (h_1, h_2, s > o)$$

are the Wald acception and rejection lines, respectively. Suppose the procedure stops when $n = N$ and $y = D$. Then the unique, unbiased estimate of p is

$$\hat{p} = \frac{k^* (N,D)}{k\ (N,D)}, \tag{5}$$

where k (N,D) and k* (N,D) are, respectively, the number of different possible paths to (N,D) from (0,0) and (1,1). In practice, the evaluation of k and k* presents a real problem. Stockman and Armitage [10] have provided the means which are practical in - but only in - the framework of a computer. They show that if acceptance occurs, the number of paths k (N,D) must be the final element in a vector which is the product $M_1 \cdot M_2 \ldots A \cdot B \cdot A \cdot B \ldots A$ of certain specified matrices $M_1, M_2, \ldots, A$, and B. Likewise, if rejection occurs, the number of paths k (N,D) is given by the first element in a vector which is also a product of matrices. The number of paths k* (N,D) is found

by straightforward modifications of the procedure used to find k (N,D). Since matrix multiplication subroutines exist in almost all computer installations, it is feasible to evaluate (5) if a computer is available.

3. Estimating the Mean θ of a Population Whose Variance V (θ) is a Finite Function of θ

Anscombe [11] gives a procedure for constructing a boundary for a sequential process in which θ is estimated with a fixed variance v^2 or with a specified coefficient of variation c. The sampling may be represented by a graph in which the cumulative sum of the observations $Z_n = \sum_I^n X_i$ is plotted against the number of observations n. Sampling continues until a specified boundary y = K (n) is crossed, where y is a general symbol for the ordinate. θ will be estimated with specified variance v^2 if the equation of the boundary is

$$\frac{1}{n} V\left(\frac{y}{n}\right) = v^2 \tag{6}$$

and with specified coefficient of variation c if the equation of the boundary is

$$(n/y^2)\, V\, (y/n) = c^2 \tag{7}$$

The estimate of θ is $\hat{\theta} = k(N)/N$, where N is the value of n for which the boundary is first crossed.

4. Estimating the Variance of a Normal Distribution

As is well known, for a fixed number ν of degrees of freedom the estimate s^2 of the variance σ^2 of a normal variate has a Type III distribution with mean σ^2 and variance

$$V(s^2) = \frac{2\,(\sigma^2)^2}{\nu}$$

Equation (6) yields the boundary

$$y^2 = \frac{\nu}{2}\, v^2\, n^3 \tag{8}$$

for a fixed variance v^2 and the boundary

$$n = \frac{2}{\nu c^2} \tag{9}$$

for a fixed coefficient of variation c. The boundary (9) is a vertical line, specifying a sample of fixed size. As before, sampling continues until the boundary is crossed.

The question naturally arises as to whether the additional work involved in the continual bookkeeping necessary in sequential procedures is worth while, or whether it might be better to pursue the simpler non-sequential procedures and work with correspondingly larger samples. If the generative process of the simulation model is relatively straightforward and not too lengthy, the additional work due to the continual bookkeeping associated with the sequential procedures is probably not worth it. On the other hand, it may well be that for an additional investment of 0.1 percent to 1 percent more computing time per sample generated, the total amount of time and money spent on the entire problem may be considerably reduced through the judicious use of sequential procedures. Since big machines are expensive, this saving may well be worth while for many simulation-type problems.

IV. OTHER METHODS FOR REDUCING THE MAGNITUDE OF A SIMULATION-TYPE PROBLEM. There are various devices for reducing the magnitude of a problem which can be applied to certain kinds of simulation-type problems. For example, if we are applying analysis of variance techniques to the results obtained by simulation procedures, it may not be feasible (even in the framework of a computer) to consider all combinations of the different factor levels. However, in some cases (depending upon whether certain interactions are negligible) it is possible to use fractional replication to reduce the magnitude of the problem to a feasible level. Again, if our problem is to find the levels of the factors which give an optimum response, we may in some cases use the method of steepest ascent [12], the sequential one-factor-at-a-time procedure of Friedman & Savage [13], and various other techniques.

It might be mentioned here that there are many simulation problems in which we are concerned with testing the homogeneity of a group of means. This is done via analysis of variance, provided the underlying assumptions are reasonably well satisfied. If the assumptions are not met, it is often possible, through the use of a transformation, to obtain transformed data which do satisfy the underlying assumptions, particularly the assumption of normality and sometimes the assumption of homogeneity of variance. However, there are situations (e.g., when the data are bimodal in nature) when "normality" cannot be achieved; and there are other cases in which the heterogeneity of variance cannot be adequately removed. While it has been shown that heterogeneity of variance usually (though not invariably) tends to affect the significance level so that too many significant results are obtained, this is often times insufficient information, particularly if we wish to known the magnitude of the change in the significance level for the general analysis of variance problem involving multiple classification. Recently considerable work has been done investigating the assumptions underlying the general analysis of variance problem involving any number of factors and levels of factors. We can only touch upon some of the results here. David and Johnson [14], [15], have developed a method for investigating the effect of nonnormality and heterogeneity of variance on tests of the general linear hypothesis. The method is based on finding the cumulants of a linear function of the two sums of squares used in the usual F test. However, application of the method is somewhat tedious, since for each F ratio it involves fitting a curve (frequently a Pearson Type IV) to the first four moments of said linear function of the two sums of squares involved and determining the critical value corresponding to the

α level of significance. Wilson [16] has proposed a distribution-free test of analysis of variance hypotheses based on a chi-square statistic for a contingency table which can be decomposed into components in much the same manner as a total sum of squares is decomposed in analysis of variance computations. This method is applicable to analysis of variance problems involving either single or multiple classification. While Wilson makes no statements concerning the power of the test, there are some indications [17] that the power of the test is low. Very recently Gurland has suggested another method which assumes normality but does not require homogeneity of variance [18]. Perhaps the most far-reaching results have been obtained by Cornfield and Tukey [19], who have developed a pigeonhole model based on average value of mean squares. The remarkable thing about this method is its flexibility and generality. With this approach, it is not necessary to postulate the type of model (components of variance, fixed effects, or mixed model), since the pigeonhole model includes all three as special cases, but without any assumptions about interactions, normality, and homogeneity of variance. Wilk and Kempthorne [20] have extended the method to include the case of randomized blocks (not treated by Tukey and Cornfield) where the variability of experimental units may be large.

REFERENCES

[1] John W. Tukey and Hale F. Trotter, "Conditional Monte Carlo for Normal Samples," Symposium on Monte Carlo Methods (1956), p. 68.

[2] Herman Kahn, "Use of Different Monte Carlo Techniques," Symposium on Monte Carlo Methods (1956), pp. 146-190.

[3] John E. Walsh, "Questionable Usefulness of Variance for Measuring Estimate Accuracy in Monte Carlo Importance Sampling Problems," Symposium on Monte Carlo Methods (1956), pp. 141-144.

[4] Symposium on Monte Carlo Methods, pp. 64-79; 80-88, 146-190.

[5] Jack Moshman, "Making Computers Cry 'Enough'!", paper presented at a meeting on the Effect of High-Speed Computing on Statistics cosponsored by the Institute of Mathematical Statistics and the American Statistical Association on September 12, 1957 in Atlantic City, N. J.

[6] Jack Moshman, "A Method for Selecting the Size of the Initial Sample in Stein's Two-Sample Procedure," submitted to the editor of the Annals of Mathematical Statistics.

[7] C. Stein, "A Two-Sample Test for a Linear Hypothesis Whose Power is Independent of the Variance," Annals of Mathematical Statistics, Vol. 16 (1945), pp. 243-258.

[8] F. J. Anscombe, "Sequential Estimation," Journal of the Royal Statistical Society, Series B, Vol. 15 (1953), pp. 1-21.

[9] M. A. Girshick, F. Mosteller, and L. J. Savage, "Unbiased Estimates for Certain Binomial Sampling Problems with Applications, "Annals of Mathematical Statistics, Vol. 17 (1946), pp. 13-23.

[10] C. M. Stockman and P. Armitage, "Some Properties of Closed Sequential Schemes," Supplement to the Journal of the Royal Statistical Society, Vol. 8 (1946), pp. 104-112.

[11] F. J. Anscombe, "Large Sample Theory of Sequential Estimation," Biometrika, Vol. 36 (1949), pp. 455-458.

[12] G. E. P. Box, "The Determination of Optimum Conditions," The Design and Analysis of Industrial Experiments (edited by O. L. Davies;1954), chapter 11.

[13] Milton J. Friedman and L. J. Savage, "Planning Experiments Seeking Maxims," Techniques of Statistical Analysis (1947), Chapter 13.

[14] F. N. David and N. L. Johnson, "A Method of Investigating the Effect of Nonnormality and Heterogeneity of Variance on Tests of the General Linear Hypothesis," Annals of Mathematical Statistics, Vol. 22 (1951), pp. 382-39

[15] F. N. David and N. L. Johnson, "Extension of a Method of Investigating the Properties of Analysis of Variance Tests to the Case of Random and Mixed Models," Annals of Mathematical Statistics, Vol. 23 (1952), pp. 594-601.

[16] Kellogg V. Wilson, "A Distribution-Free Test of Analysis of Variance Hypotheses," Psychological Bulletin, Vol. 53 (1956). pp. 96-101.

[17] Quinn McNemar, "On Wilson's Distribution-Free Test of Analysis of Variance Hypotheses," Psychological Bulletin, Vol. 54 (1957) pp. 361-362.

[18] John Gurland, "Testing Homogeneity of Means in the Presence of Heterogeneity of Variance," paper presented at the meeting of the Institute of Mathematical Statistics on September 13, 1957 in Atlantic City, New Jersey.

[19] Jerome Cornfield and John Tukey, "Average Values of Mean Squares in Factorials," Annals of Mathematical Statistics, Vol. 27 (1956), pp. 907-949.

[20] M. B. Wilk and Oscar Kempthorne, "Some Aspects of the Analysis of Factorial Experiments in a Completely Randomized Design," Annals of Mathematical Statistics, Vol. 27 (1956), pp. 950-985.

PROBLEMS IN ANALYSIS OF ELECTRON TUBE EXPERIMENTS

M. H. Zinn
U. S. Army Signal Engineering Laboratories

In attempting to apply the technique of experimental design to the investigation of electron tube characteristics, one is faced by the fact that the devices being tested are complex entities, each part of which has gone through many processes and hands. Since any one of these processes can affect the end product, the experimenter does not have under his control all of the factors required to reduce his error (or the variance due to unknown factors) to a minimum. Even if all of the factors could be controlled to the extent that a single set of conditions of manufacture could be exactly repeated, the results of the experiment could not necessarily be extended to cover all possible conditions in the general production of the tube type by other sources. A further complication arises when the experiment is concerned with determining the effects of extended operation under imposed levels of operating conditions on specific tube characteristics. In this type of experiment one usually finds that the tubes react to the imposed conditions in such a way that the levels of the real variables are dispersed. Control of the real test variable under these conditions is either extremely difficult or extremely costly. The general consequences of this lack of complete control over the test variables are that experiments involving electron tubes must be performed using moderately large samples in order to obtain significant results. Once one has paid for the samples, test equipment, operation time, and test time involved in the basic statistical design, it will usually be profitable to perform small experiments outside the basic design to assist in assigning causes to any significant differenoes which may show up in the experiment or to reduce the general experimental error

As a means of illustrating this general type of problem in setting up a design of experiment for electron tubes and some of the questions which may arise during the analysis period, I would like to discuss a particular experiment, initiated by our Laboratories, which is being performed by Briggs Associates, Inc. of Norristown, Pennsylvania.

One of the detrimental factors that occurs in oxide coated cathode types of tubes is the formation of a resistive interface layer between the base metal and the coating material. The appearance of this resistive layer during the life of the tube results in a degradation of tube performance. Research studies have indicated that the layer is formed at the interface due to a chemical reaction between the coating material and impurities in the nicked base metal forming a compound such as Ba_2SiO_4 (Barium orthosilicate) Impurities other than silicon, such as magnesium, manganese, etc., may also form compounds, but it is the consensus of opinion that the orthosilicate is responsible for the high resistance interface layers. At this point one might say, "Well, get rid of the silicon or other impurities and you have solved all your problems." This is somewhat easier said than done since the activation of the cathode to produce the required emission is dependent upon the presence of impurities to a large extent. In addition, it has been found that the same base materials when used by different manufacturers do not necessarily yield the same results, which indicates that processing or the exact nature of the coating may be involved. Finally, it is known that the interface resistance is usually low when the tubes are first produced and that the interface grows during the life of the tube at a rate dependent upon the conditions of operation. In setting up an

experiment to determine factors affecting the growth of interface resistance it was necessary, therefore, to consider base metals, manufacturers, and combinations of operating conditions over a life period long enough to permit growth of the interface layer.

The experiment was set up using a factorial design covering

3 manufacturers

4 base metals

3 levels of filament voltage

3 levels of plate current

The complete factorial design thus calls for 3 times 4 times 3 times 3, or 108 individual cells. The number of tubes to be tested in each cell was determinined using the equation

$$\sqrt{n} = (\mu_{1-\beta} + \mu_{1-\alpha/2})\ \sigma/\delta$$

n = number of tubes in a group capable of being averaged

β = .05 error of the second kind

α = .05 error of the first kind

σ = expected standard deviation of the difference between means of averageable groups

δ = minimum desired detectable difference in means

$\mu_{1-\beta}$ and $\mu_{1-\alpha/2}$ are unit normal deviates corresponding to the values of $1-\beta$ and $1-\alpha/2$

Although it was planned to use the analysis of variance to examine the null hypotheses of no row or column effects for several tube characteristics, since interface layer resistance was the basic subject of the analysis, the minimum sample size was established based on the expected distribution of this one characteristic. Based on a limited amount of data, it was assumed that interface resistance would have a log normal distribution with a standard deviation of .358 on a logarithmic base for any one homogeneous group. This resulted in the use of a value of $.506\sqrt{2}\,1.358$ for α in the equation as the standard deviation of two tube means. A minimum detectable difference of 0.146 was established which corresponds to a mean 1.5 times that of another group on an arithmetic base. For the values of .05 chosen for α and β this equation results in a minimum number equal to 107 tubes for each group which predictable included no interactions. If we assumed that interactions could be present between each individual cell, 108 times 107, or 11,556 tubes would be required for the experiment.

It was decided that we were particularly interested in manufacturer-alloy interactions and that the possibility existed that a regression surface could be plotted for the two numerical factors in the experiment--filament voltage and plate current. Based on these expectations, the experiment was initiated with 117 tubes in each alloy-manufacturer group. The extra ten tubes represent a small margin of safety to allow for some catastrophic failures over the 5000-hour life period. An additional margin was provided by choosing as a test vehicle a twin triode, but a means of using this margin of safety was not devised at the onset. At any rate, a total of 234 triode units in 117 envelopes were to be tested for each alloy-manufacturer group (1404 tubes--2808 sections)

The experiment has proceeded to the point that 63 tubes in each alloy-manufacturer group have been tested for 5000 hours and the remaining 54 tubes in each group are nearing the mid-point in life. Side experiments have been performed--some successfully and some unsuccessfully. On the unsuccessful side of the ledger was an attempt to make quantitative measurements of cathode temperature. This information was required since we can expect a wide variation of cathode temperature at a specific filament voltage within any one manufacturers' tubes as well as variations in the mean temperature between manufacturers' tubes. Since the cathode temperature is the basic variable which one attempts to control through the applied levels of filament voltage, the data would have been of extreme value in reducing variance in the test results. The problem is further complicated by the fact that a major reaction of the tubes with the imposed conditions, as discussed briefly in the introduction is anticipated in this area of cathode temperature. This reaction occurs in the following manner:

1) The rate of growth of interface is a function of the cathode temperature.

2) The formation of the interface layer modifies the power radiated by the cathode due to a change in spectral emissivity.

3) As a result of the change in radiated power, the cathode temperature changes at the constant levels of filament voltage.

The level of cathode temperature will, therefore, be changing continuously during the course of the experiment, again contributing additional variance of the test results. A general treatment of this problem of variable basic levels of test in statistical designs has not been made and is worthy of some attention by the statistician.

On the positive side of the ledger, as far as side experiments are concerned, are the results of spectrographical analyses of the alloys used for the cathode base metals. Analyses were made on the metal as it came from the original supplier of cathode sleeves and on samples taken from control tubes taken from each alloy-manufacturer group before tubes were placed on life test. Additional tests will be made on samples taken from the groups which have completed 5000 hours. These analyses will be used in an attempt to assign causes for significant differences due to alloy-manufacturer interactions and alloy-manufacturer condition interactions if these significant differences are found in the statistical analysis.

The analysis itself will not be undertaken until all the data have been collected. In the meantime, however, preliminary tests of the data can be made in order to guide the final analysis. Analysis of variance techniques on small segments of date, such as within one alloy-manufacturer group, which can be qualitatively compared with results on another alloy-manufacturer group, indicate that some of the interactions that were assumed to be absent are indeed present. These preliminary results indicate that it is essential to find a method of treating our twin triodes. That is, for n number of tubes and 2n number of triode sections under what conditions can we treat the problem as though we had 2n tubes under test?

We have proposed to test for the independence of the triode sections in the following manner:

Assume we were drawing independent samples of two tubes at a time from a population which has a homogeneous normal distribution for a particular test characteristic. In this case the variance of the differences between paired readings is related to the variance of the population

$$\sigma_d^2 = \frac{1.18\ \sigma^2}{2} \quad \left(\begin{array}{c} \text{1.18 obtained from} \\ R + \bar{R} \text{ data for} \\ DL_1 \text{ and } DU_1 \text{ and } A_2 \end{array} \right)$$

If we can obtain unbiased estimates of the variance of the paired observations and the variance of the population, we could perform an F test to test as a null hypotheses that there is no reason to doubt that the individual readings are independent. This will be done by calculating S_d^2 the variance of the difference between triode sections, and using S_L^2 and S_R^2 the variance of the left and righthand sections, respectively, to obtain a pooled estimate $\frac{S_L^2 + S_R^2}{2}$ of σ^2.

The F test would then take the form

$$F_{a/2}\ (n-1,\ 2n-2) < \frac{3.39 S_d^2}{S_L^2 + S_R^2} < F_{1-a/2}\ (n-1,\ 2n-2)$$

The test will result in one of three possibilities:

F within critical limits: accept null hypotheses of independence of triode sections and treat as though 2n samples are available.

If F < then lower critical limit reject null hypotheses of independence. Since the triode sections are not independent, treat data as though only n samples are available.

If F < then upper critical limit reject null hypotheses of independent samples from homogeneous population. Examine data further for significant difference in means of left and righthand sections and treat

data as though two separate samples of n each are present in the data.

It is possible that the entire technique handled separately from the analysis of variance can be avoided. Comments on the technique or suggestions of alternative approaches will be appreciated.

Another problem in the analysis which requires some treatment is the procedure for handling time on test as a variable. Several alternatives are available. The simplest approach is to plot either individual tube readings as a function of time or to consolidate these plots into cell groups and show average value and the dispersion around the average. A second approach would be to use the analysis of variance technique over all the time periods for which measurements were made, treating time as a variable in the same manner as metals, manufacturers, or operating conditions. A third approach, and the one which is proposed for use in the analysis, would be to perform independent analyses of variance on the data collect at separate periods of time. Once one has determined whether significant effects are present at a particular point in time and where the effects lie, take an avera over all groups capable of being averaged and show confidence limits for the avera If this is done at all periods of time for which readings have been made then a connected plot can be formed. A test of significant differences between means at one point in time could be made with the next previous reading to prove that any slope in the connecting line is justified.

SOME PROBLEMS ENCOUNTERED IN THE EVALUATION OF EROSION IN CANNON BORES

P. J. Loatman
Watervliet Arsenal

INTRODUCTION. Before discussing two tests which are in the predesign stage, let us consider, very briefly, some general aspects of erosion. It is not surprising that each time a round is fired a small amount of metal is removed from the bore; as firing progresses the dimensions of the bore gradually change. The wear involved is not quite so simple as that occurring when two pieces of material are rubbed one against the other. Many things enter into what is termed erosion; the literature usually treats these under the general headings of thermal, chemical and mechanical factors. They do not act independently of each other, the process being characterized by the simultaneous interaction of all factors. Appreciable wear is not general throughout the tube but is localized in the region of the origin of rifling. This is termed origin erosion and is usually the type considered of first importance. Muzzle velocities of the order of 2500 ft/sec and higher induce another type of localized wear in a region near the muzzle. We shall, however, confine ourselves to a very brief description of origin erosion.

Upon ignition of the charge a steep temperature gradient develops in the wall of the tube. The heat generated causes a softening of the inner steel layers; the accompanying high temperatures promote chemical reactions between the steel and the products of the burning powder gases. Iron oxide and iron carbide are formed, and a so-called white layer may develop. The combined effect of the thermal and chemical factors is such that the bore interface is predisposed to erosive action by the projectile. The forward motion of the projectile sets up additional stresses and removes steel from the region of the origin of rifling. Gas "blow-by" may also accelerate the erosion. All of these various events occur very rapidly, being something of the order of several milliseconds.

In common with most wear phenomena the rate of erosion decreases when compared to that obtaining initially. Concurrent with the increase in the dimensions of the bore, the muzzle velocity decreases (plate No. 2).* The fall off in velocity decreases with increasing round number. It is evident that muzzle velocity may serve as an index of wear.

The type of round is one of the chief factors determining the amount of wear. The upper portion of Plate No. 3 shows that generally AP rounds induce different wear than HE rounds. If the type of round remains the same but the amount of propellant is varied a different rate of wear obtains (lower portion of Plate No.3).

The amount of wear decreases as we go from the origin towards the muzzle (upper portion Plate No. 4). A cone of wear develops in the origin of rifling region.

Neither origin nor muzzle erosion is symmetrical. The latter, however, exhibits the greater asymmetry; one type is shown in the lower portion of Plate No. 4.

---

*Plates can be found at the end of the article.

Plate No. 5 illustrates the effect of three modifications in the forcing cone for three 120mm. guns. A correlation existed for the data of curve A; it is merely a plot of the regression equation. B and C, however, did not correlate. These latter curves were "drawn by eye". The purpose here is to show that we do not always obtain the nice, simple curves of the previous plates.

The two tests to be discussed could be considered important merely from the viewpoint of cost. Considering only the cost of the ammunition and testing, the first will approximate \$100,000 and the second \$500,000. Since costs are so high for two relatively simple tests, it behooves us to conduct them so as to achieve the maximum amount of information.

Three or four agencies are involved in these tests; this fact alone complicates things. It is only natural that each agency should seek to have its own interests served. In what follows I have outlined alternate tests to those proposed by other groups. The prime purpose of my proposals is to generate discussion at this session and arrive at what we might term the best possible solution.

TEST NO. 1 - MODIFIED GUN. Verification of predictions on the performance of a modified gun is the aim of the first test. By relatively inexpensive means it is possible to increase the effectiveness of an existing gun. Table No. 1 outlines a test sequence which will probably be used.

Pressures, muzzle velocities and wear measurements will be taken during the test. Wear measurements will not be taken after every round, but at the best, after every 10 rounds. It may well be that the wear will be measured only at the beginning and end of test. Pullover gage readings could be taken at frequent intervals at relatively little cost. While these are not so accurate as star gage readings they would provide much useful wear data.

The sequence of Table No. 1 enables us to detect if a significant difference exists between guns. This is true, however, only for half of the test. Firing all of the HEP rounds in Gun No. 2 in the fourth test precludes any valid comparison of worn tube accuracy at the end of test. In Plate No. 3 it was noted that a different rate of wear must be expected for HE rounds than that obtaining for AP rounds. The expectancy for the HEP rounds is that there will be appreciably less wear than for the AP round.

The accuracy tests for both guns involves ammunition conditioned at three different temperatures. It is anticipated that the time interval between rounds may be as much as 45 minutes. The guns will be at ambient temperature for the firings. As given here the wear is always biased in favor of the 70°F rounds over the -40°F and 125°F rounds. If calibration rounds were to be fired at different times throughout the sequence, then a correction could be made for wear. But calibration rounds are not available. Further, mount M2 is used in each case and it is radically different from M1. In service the gun will be used on M1, hence it is important to know the accuracy on this mount. It is understood that those conducting the test will use past experience to estimate the performance of the guns on M1. Since they have never tested this type of gun before, how reliable an estimate can be made from past experience?

Test No. 5, Gun No. 1 involves four webs for each of two types of propellant. It is desired to determine the best web for each propellant. As scheduled a wear bias will exist in favor of the first web tested. As was the case for the accuracy test, here too there is no question of using calibration rounds.

Table No. 2 outlines an alternate procedure. Here we have "balanced" the test, i.e., each gun fires the same number and type of ammunition. (The rate of wear for Slugs and APDS rounds is substantially the same). It is believed that much more information will be gained from the accuracy phase of the test if it is conducted as a simple four factor factorial experiment. Three of the factors are at two levels and the temperature is at three levels. Since the accuracy test is unconfounded each round may be assigned a number and by randomization the firing sequence determined. We can determine the four main effects and six two-factor interactions. The four three-factor interactions available are not of interest. Due to the number of rounds and the expected large time interval between rounds, it will probably be necessary to confound. In this event randomize the order of the blocks and then randomize the order of the rounds within the block. It may be better to restrict the randomization so that we always fire (in sequence) two rounds which are at the same temperature. This will enable us to determine the round to round variation. In view of the anticipated large time interval between rounds, however, it is doubtful that any great advantage will accrue from this procedure.

Traditionally, accuracy is determined by firing in groups of five or ten. Randomization will necessitate identifying each round on the target. This presents no great problem for usually men are stationed down range to mark each hit; likewise the round can be identified from the gun site by the use of a telescope. Tests numbers 5 and 7 are merely three factor factorials with randomization.

TEST NO. 2 - EXPERIMENTAL GUN. The second test aims at establishing the overall reliability for an experimental gun. One of the four guns available for test will be modified throughout its testing. The remaining three guns, however, are to be tested as outlined in Table No. 3. To compare the results of Gun No. 1 and 2 it is necessary to assume that the difference between guns is negligible. Prototypes of the present gun have exhibited rather wide variations from gun to gun in their ballistic behavior.

Tests on Gun No. 3 involve three lower temperatures. It is feared that the gun may be destroyed at -65°F. It is desired, therefore, to gradually approach the lower level. A very marked temperature effect is expected. It is believed that a temperature of -20°F. will not damage the gun. It is proposed that only two temperature levels be considered for the gun and ammunition. Table No. 4 outlines one possible arrangement for combining the tests of Table No. 3 into a five factor factorial experiment at two levels. Depending upon the circumstances at the time, it might be more feasible to run the experiment in eight blocks with four observations per block. It will then be necessary to confound two of the two-factor interactions. Eight other guns are involved in this test but they are under the cognizance of other agencies. Their test sequence is essentially the same as that outlined in Table No. 3.

The actual testing program of the tests discussed here will probably extend over a three or four month period. The firings would not take that long if they were conducted on a continuous basis. Various factors will compel a halting of the firings from time to time. In view of this the possibilities of fractional replication seem very attractive.

REPRESENTATIVE OF SOME GUNS USING FIXED AMMUNITION

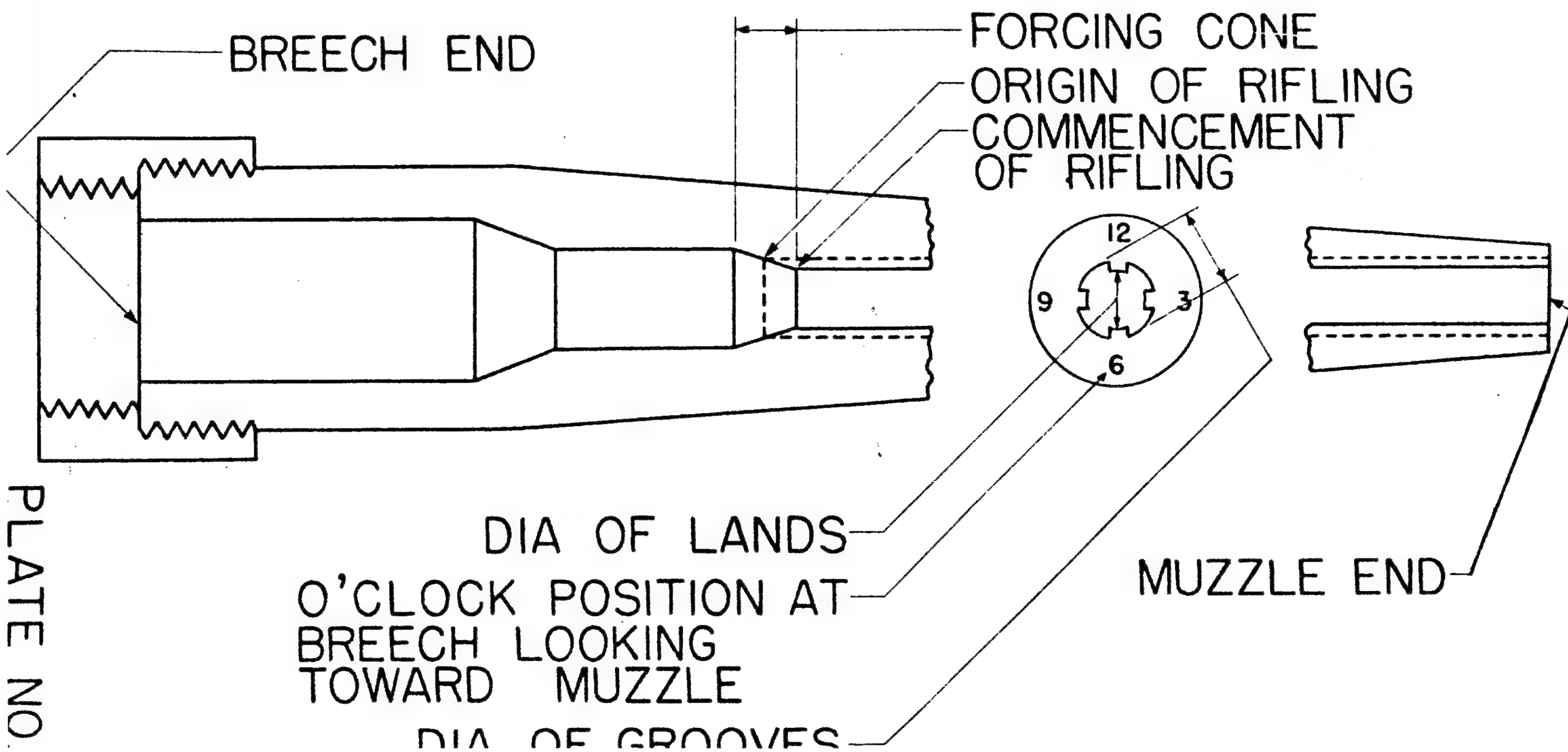
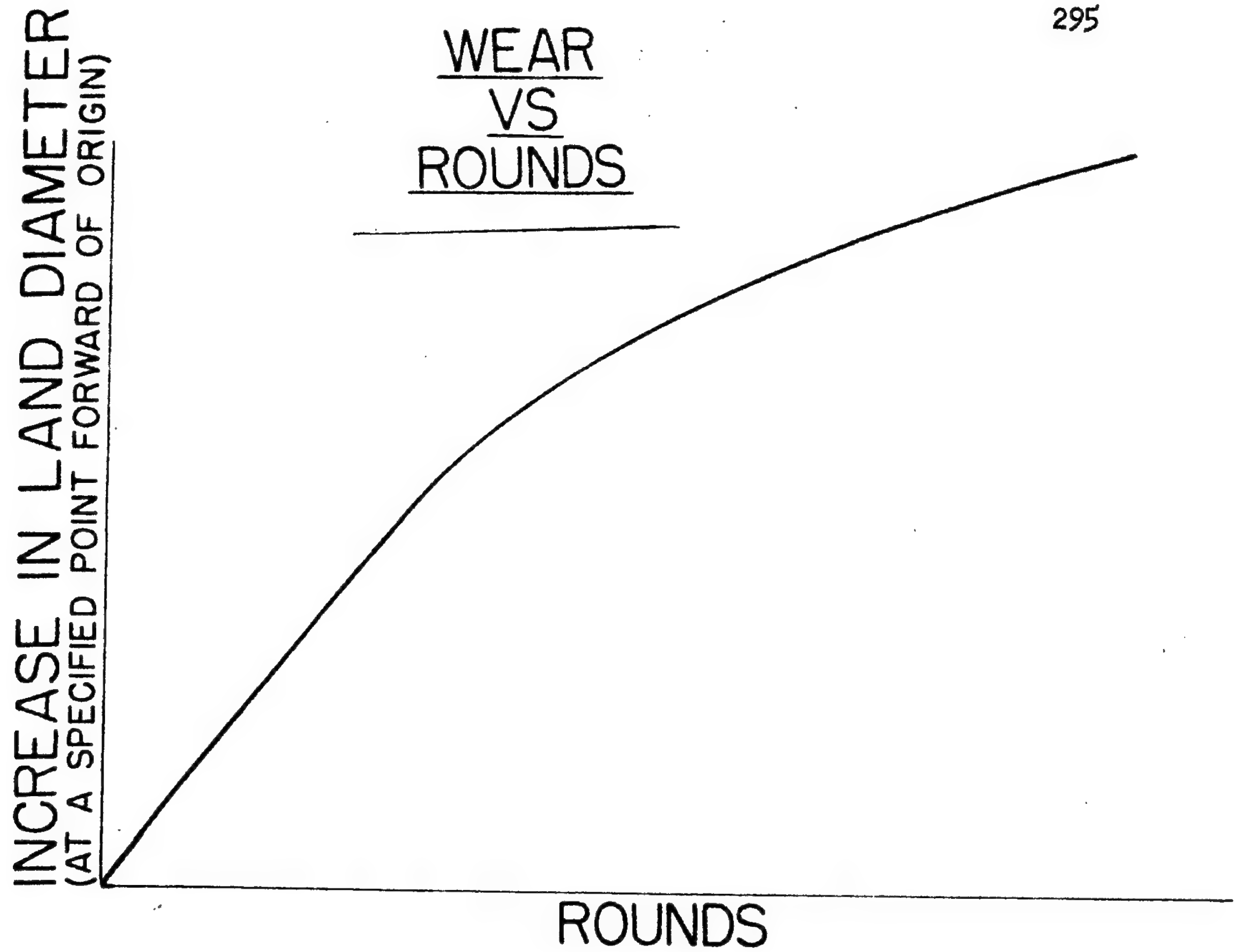
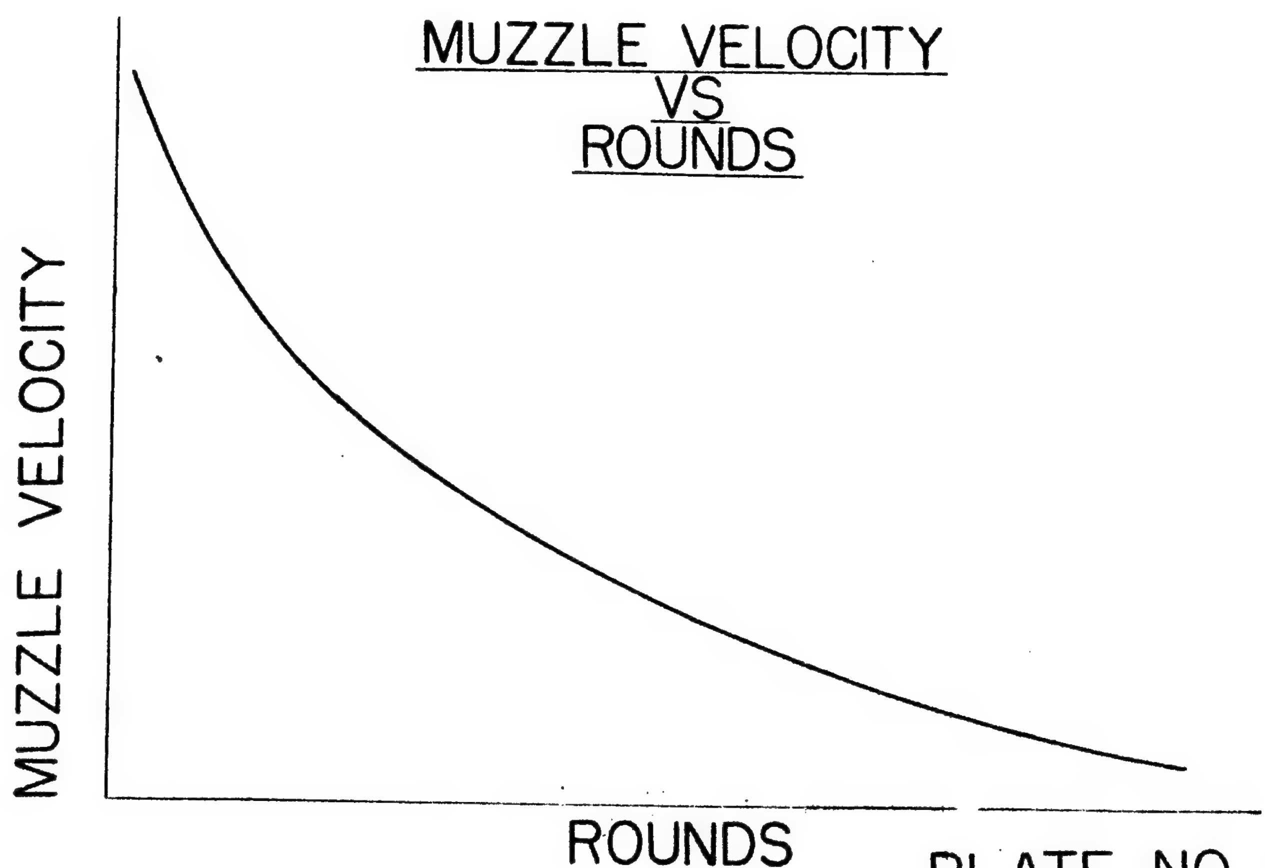


PLATE NO.



WEAR
VS
ROUNDS
INCREASE IN LAND DIAMETER
(AT A SPECIFIED POINT FORWARD OF ORIGIN)
ROUNDS



MUZZLE VELOCITY
VS
ROUNDS
MUZZLE VELOCITY
ROUNDS

PLATE NO.

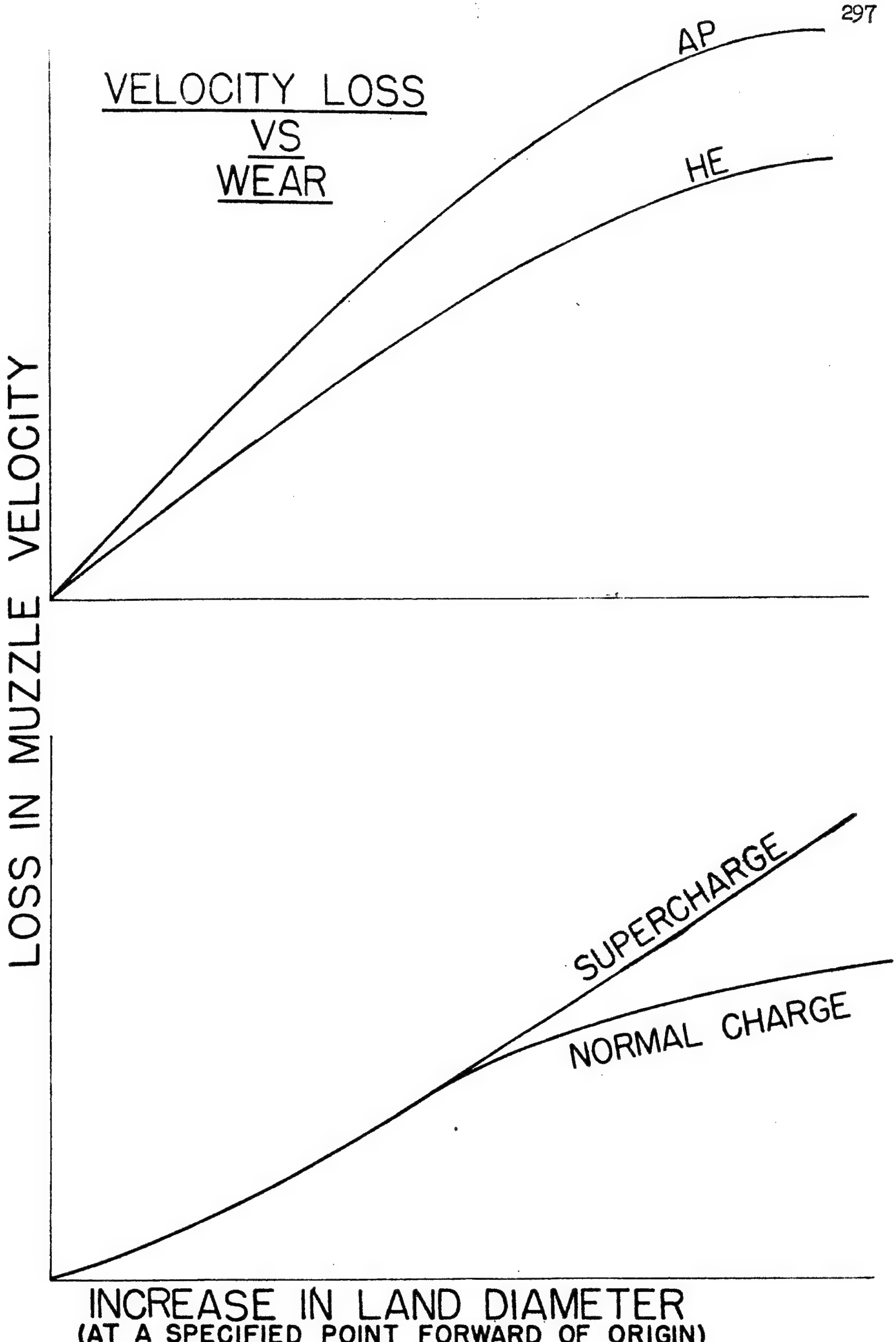
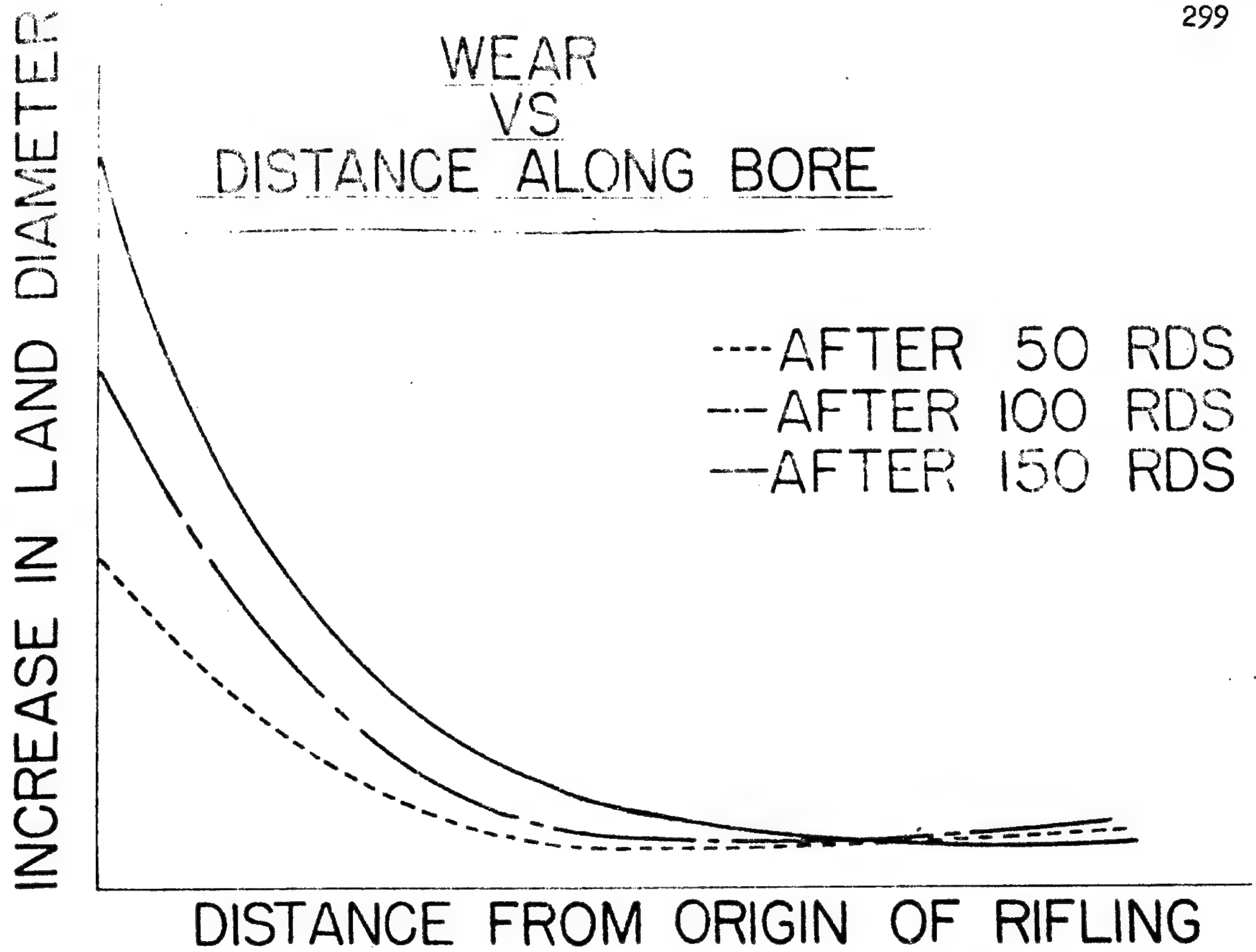

VELOCITY LOSS
VS
WEAR
AP
HE
LOSS IN MUZZLE VELOCITY
SUPERCHARGE
NORMAL CHARGE
INCREASE IN LAND DIAMETER
(AT A SPECIFIED POINT FORWARD OF ORIGIN)

PLATE NO.


WEAR
VS
DISTANCE ALONG BORE
----AFTER 50 RDS
-·-AFTER 100 RDS
—AFTER 150 RDS
INCREASE IN LAND DIAMETER
DISTANCE FROM ORIGIN OF RIFLING

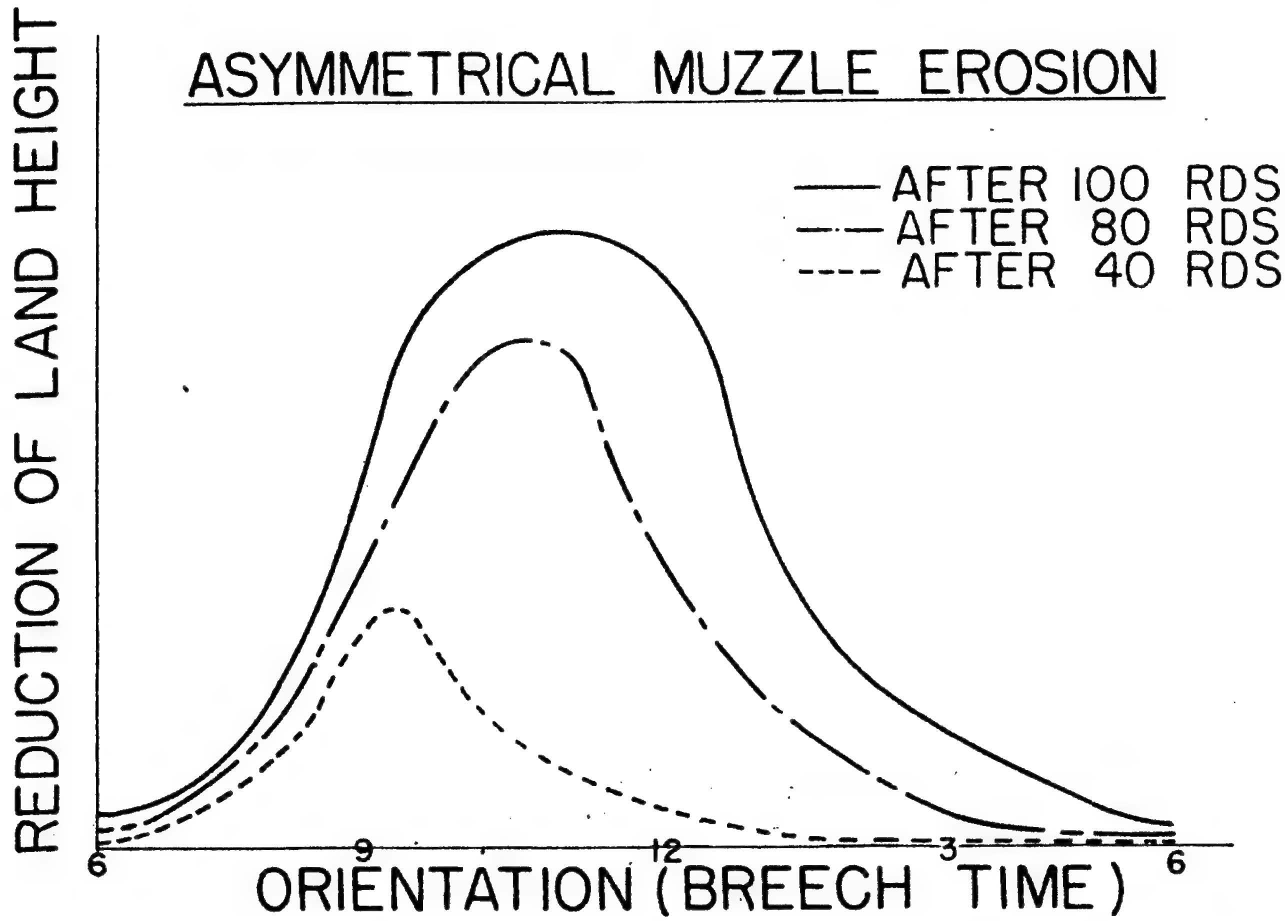

ASYMMETRICAL MUZZLE EROSION
—— AFTER 100 RDS
-·- AFTER 80 RDS
---- AFTER 40 RDS
REDUCTION OF LAND HEIGHT
6
9
12
3
6
ORIENTATION (BREECH TIME)

PLATE NO

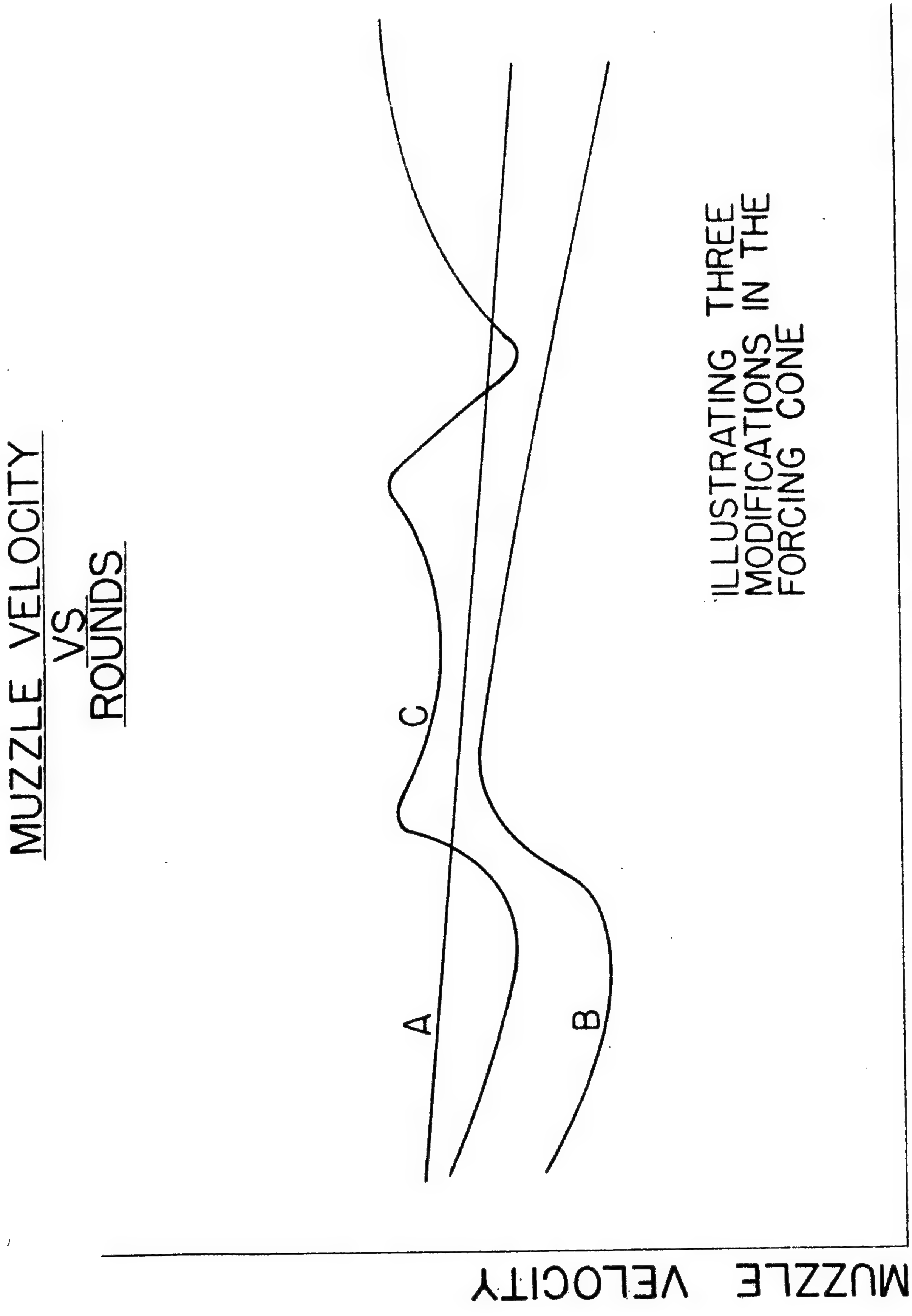

MUZZLE VELOCITY
VS
ROUNDS
ILLUSTRATING THREE
MODIFICATIONS IN THE
FORCING CONE
A
B
C
ROUNDS
MUZZLE VELOCITY

PLATE NO.5

TABLE NO.1
COMBINED AGENCIES SEQUENCE FOR CONDUCTING TEST

	GUN NO.1				GUN NO. 2		
TEST NO.	MOUNT	DESCRIPTION	RDS.	TEST NO.	MOUNT	DESCRIPTION	RDS.
1	M1	EVALUATE PROPELLANT	5 SLUGS	1	M2	ACCURACY Range 1 70 F -40 125F Range 2 70 F -40 F 125 F	 10 APDS 10 APDS 10 APDS 10 APDS 10 APDS 10 APDS
2	M2	ACCURACY Range 1 70 F -40 F 125 F Range 2 70 F -40 F 125 F	 10 APDS 10 APDS 10 APDS 10 APDS 10 APDS 10 APDS	2	M1	PROOF RECOIL MECHANISM	5 SLUGS
3	M2	BALLISTIC LIMIT	60 APDS	3	M1	ACCURACY & JUMP	40 APDS
4	M2	SP TEST	20 APDS	4	M2	HEP TEST	48 HEP
5	M1	ESTABLISH CHARGE Propellant 1 4 Webs Best Web Propellant 2 4 Webs Best Web	50 SLUGS	5	M1	MISCELLANEOUS	40 SLUGS
6	M2	WORN TUBE ACCURACY	20 APDS	6	M2	WORN TUBE ACCURACY	20 APDS

TABLE NO. 2

PROPOSED SEQUENCE
FOR CONDUCTING TEST

TEST NO. 1
EVALUATE PROPELLANT
5 SLUGS EACH GUN

TEST NO. 2
ACCURACY
60 APDS EACH GUN
RANDOMIZE

Gun 1				Gun 2			
Mount 1		Mount 2		Mount 1		Mount 2	
Range 1	Range 2	Range 1	Range 2	Range 1	Range 2	Range 1	Range 2
T1 T2 T3	T1 T2 T3	T1 T2 T3	T1 T2 T3	T1 T2 T3	T1 T2 T3	T1 T2 T3	T1 T2 T3

TEST NO. 3

PART A
Gun 1, Mount 1
BALLISTIC LIMIT
60 APDS

PART B
Gun 2, Mount 2
ACCURACY & JUMP
40 APDS
MISCELLANEOUS
Phase 1
20 SLUGS

TABLE NO. 2 CONTINUED

TEST NO. 4
HEP ROUND TEST
Panel Fragmentation
2 HEP Each Gun
Terminal Ballistics
12 HEP Each Gun
Accuracy + Terminal Ballistics
10 HEP Each Gun

TEST NO. 5
ESTABLISH CHARGE
40 SLUGS EACH GUN
RANDOMIZE

Gun 1		Gun 2	
Propellant 1	Propellant 2	Propellant 1	Propellant 2
W1 W2 W3 W4	W1 W2 W3 W4	W1 W2 W3 W4	W1 W2 W3 W4

5 SLUGS BEST WEB, EACH GUN

TEST NO. 6

PART A
Gun 1, Mount 1
SP TEST
20 APDS

PART B
Gun 2, Mount 2
MISCELLANEOUS, Phase 2
20 SLUGS

TEST NO. 7
WORN TUBE ACCURACY
20 APDS EACH GUN
RANDOMIZE

Gun 1				Gun 2			
Mount 1		Mount 2		Mount 1		Mount 2	
Range 1	Range 2	Range 1	Range 2	Range 1	Range 2	Range 1	Range 2

TABLE NO. 3

PRESENT SEQUENCE FOR TEST

GUN NO. 1: ALL ROUNDS FIRED AT 100% CHARGE (100 AP RDS.). DETERMINE a) GUN FREQUENCY b) MAGNITUDE OF MUZZLE WHIP c) MUZZLE VELOCITY d) PRESSURE e) BREECH STRAIN f) P. E.

GUN NO. 2: SAME AS FOR GUN NO. 1 EXCEPT ALL ROUNDS FIRED AT 115% CHARGE

GUN NO. 3: SAME AS FOR GUN NO. 1 EXCEPT FIRE HE ROUNDS. TUBE TO BE DRILLED IN 6 LOCATIONS DOWN LENGTH.

TEMPERATURE COMBINATIONS

DEGREES F		DEGREES F	
Gun Ambient	Rd. Ambient	Gun	Rd.
		-40	+70
0	+70	-40	-40
0	0	65	+70
-20	+70	-65	-65
-20	-20		

TABLE NO. 4

PROPOSED TEST

FOR STATISTICAL PURPOSES LET

A:	GUN NO.1	a:	GUN NO. 2
B:	100% CHARGE	b:	115% CHARGE
C:	AP ROUND	c:	HE ROUND
D:	GUN TEMP. +70° F	d:	GUN TEMP. -20°F
E:	AMMO. TEMP. +70°F	e.	AMMO. TEMP. -20°F

BLOCK 1	(1)	de	bc	bcde	abe	abd	ace	acd
BLOCK 2	e	d	bce	bcd	ab	abde	ac	acde
BLOCK 3	b	bde	c	cde	ae	ad	abce	abcd
BLOCK 4	a	ade	abc	abcde	be	bd	ce	cd

ABC, ADE, & BCDE are confounded

DETERMINING DURABILITY OF TEXTILE FABRICS BY MEANS OF CONTROLLED FIELD TESTING

John W. Griswold
Quartermaster Research and Engineering Field Evaluation Agency

After newly developed items have been passed by laboratory tests a very important question still remains to be answered. How will the items react with respect to the acceptability standards and wear imposed on them by the consumers? To answer this question on the items of Quartermaster Corps responsibility is the mission of the Quartermaster Research and Engineering Field Evaluation Agency at Fort Lee, Va. At this Agency both controlled and accelerated and normal use of testing of newly developed or improved items are performed under actual or simulated field conditions. Items tested include aerial delivery equipment, individual clothing and equipage, organizational equipment, and class I supplies, including troop acceptability of rations.

This paper concerns a problem in the accelerated durability testing of textile fabrics. Or, more specifically, the weighting system used to score the wear damage evident on fabrics after subjecting them to repeated traversals of an accelerated wear course and the methods of using it for determining the durability of textile fabrics will serve to emphasize the importance of this problem.

Shortly after the entry of the United States into World War II a course was developed at Fort Lee, Virginia for use in controlled accelerated durability tests of textiles. The word accelerated is used since it has been found through research studies that one traversal of this course corresponds to approximately one week of normal wear.

Throughout the course artificial abrasive surfaces have been avoided. Emphasis has been placed on surfaces and obstacles that will minimize the occurrence of accidental damage and produce the type of wear damage resulting from field wear. The course, as constructed, requires the normal "run-and-hit-the-dirt" type of infiltration procedure rather than acrobatics. Accordingly, the effects upon articles worn are more nearly accurate approximations of actual field wear damage than are the effects of accelerated laboratory tests.

The course consists of 29 obstacles and is approximately 1320 feet long. The repeated impact as the soldiers hit the dirt time and again, produces failures in the clothing at the knees, elbows, and body front. As the man falls, there also is some strain on shoulders, armpits, crotch, and legs. As he continues through the course other obstacles produce additional types of wear.

This is the course used for the testing of cotton and cotton-synthetic blend fabrics. For the testing of wool and wool blend fabrics several of the more damaging obstacles are eliminated. Thus, these less durable fabrics can be worn for an increased number of cycles on the course to obtain a better discrimination of minor differences.

Over the years continuous study of the testing methods used has resulted in many improvements. For example, it was found that using only trousers gave equally efficient results as the use of both shirt and trousers. Also, as another step toward increasing the number of garments that could be made from the experimental fabrics to be tested was the determination that using trousers with only the front made from the experimental fabrics resulted in no loss in efficiency. A specially designed trouser with pockets, fly, buttons and seams eliminated or relocated reduced the possibility of snags, and wear of an accidental nature. And to reduce the number of sizes of test garments required and the wide variability in wear patterns of test subjects, measures were established for obtaining modal groups of test subjects.

This modal group is obtained using both physiological and wear pattern screening. Only those test subjects in good physical condition, under 28 years of age, with weight between 125 and 165 pounds and height between 5-6" and 5-11" are considered for selection. Each member of the selected group, which usually numbers 65 test subjects, is issued two standard fabric trousers which they alternate daily in traversing the fabric course for 6 traversals per day until each trouser has received 24 traversals of wear. After each day's traversals the garments are laundered, inspected, and the wear damage charted by type, size, location and day of occurrence on a charting sheet. Wear scores for the individual garments are obtained using a system of weights for failure types by degree of failure. Analysis of these scores makes possible the elimination of test subjects whose wear is inconsistent and unusually severe as compared with the over-all group mean. This screening run also serves as a conditioning and orientation run as nearly as possible in an identical manner.

During the wear test two traversals of the fabric course constitute one cycle. At the end of each cycle of wear the garments are laundered, inspected and charted as explained for the screening phase. The design used in a wear test itself is a randomized block design. With four fabrics to be tested, forty test subjects selected by a screening run, are broken down into four sub-groups of ten test subjects. Each test subject is issued one pair of trousers in each of the four fabric types which he wears on the course, alternating fabric types after each cycle of wear. The order of wear of the fabric types is randomized between sub-groups, so that each fabric type is represented during each wear cycle. Three cycles, six traversals, of the Fabric Course are completed per day by each test subject.

At the end of the 10th cycle of wear, damage shown on the individual garment wear charts is scored to obtain wear scores by cycle.

The simple randomized block design used for Fabric Course tests has been found most practical since it is easy for the test observers to administer, allows measurement of the relatively large variation between the wear scores of test subjects, randomizes the effects of weather conditions, and is less affected by loss of test subjects than more complex designs which are used in some of our other testing at the Agency.

Many times, due to the large number of diverse types of field tests to be run at a particular time, test subjects are at a premium. At such times if the test involves comparing only two or three fabrics, each test subject is issued two garments of each type for wear on the Fabric

Course. The standard error per unit as a percent of the mean of 31% found for tests so designed compares favorably with the 25%-30% found for the usual design used.

As mentioned, earlier tests have been run to measure the correlation between Fabric Course wear and normal wear. This measurement makes possible an approximation of the amount of normal field wear produced on fabrics by our controlled accelerated tests.

Although very useful results are being obtained from the Fabric Course tests as presently conducted, continuing investigations are carried on by the Agency to improve the test methods used. These investigations have included studies to determine the consistency of charters; the length of training period required to develop a prescribed level of consistency within charters; and the efficiency of measuring wear damage using light transmission through the fabrics or by weight loss of garments worn. One investigation we have made recently concerns the validity of including wear areas in the wear scores. The fabric testing that has already been done has resulted in the development of considerable more durable fabrics and fabrics which it is more difficult to detect evidence of wear areas through visual inspection. Holes, tears and frays, of course, when they occur are clearly and easily identified. Logically, wear areas in a particular location would precede the development of holes. But due to the difficulty of detection of wear areas on some types of fabrics many holes are charted with no wear area shown. Also, on several recent tests re-charting of worn garments has been done and inconsistencies in the charting of wear areas has been found. Therefore, until a more accurate means than visual detection of wear areas is available, it may be necessary to limit the charting to holes, tears and frays. One possible method for more accurate detection of wear damage is through the use of X-Ray or similar device and studies are being made in this area. Much thought and effort have also been directed toward the improvement of the weighting system used in determining the garment wear scores; however, it is felt that further refinements are possible.

When the Fabric Course was first developed, the weights used were those that had been set up for use on a study of salvaged clothing obtained from Posts, Camps and Stations in the Continental United States. It was a system of linear weights ranging from 1 to 5 depending on the maximum diameter of the damage. Since this salvage study was run at a time when the rapidly expanding Army was taxing the production capacity of new clothing, the repairability of the clothing was the primary factor considered in the weighting system used. With only minor modification this system was used on accelerated wear tests run prior to 1945, although it was recognized by some, prior to this date, that the wear scores should reflect more the state of deterioration of the fabrics than their repairability. Around 1945 another salvage study of clothing was conducted, and it was decided to use some of these garments for an investigation into the wear score weights. Two hundred garments were randomly selected for this study, excluding those garments salvaged for burned areas, rips or other damage of an accidental nature.

The wear damage of these 200 garments was charted and the charts randomly broken down into 10 sub-groups of 20 charts each. A group of 10 men, skilled in salvage and charting work, arranged the wear charts ascending order of wear and then assigned preliminary wear scores based on their judgment and experience. The charts were scored using a variety of different weight systems. The weight system, whose derived score correlated most closely with the scores of the experts, was the one selected for a new scoring system - the one presently in use.

Fabrics selected as most durable on the Fabric Course have been standardized for troop issue. The increased durability shown by these fabrics over the standard they replaced has furnished evidence of the ability of the present scoring system to rank fabrics as a wear "end point" on the Fabric Course. Questions have been raised, however, as to whether the wear score at this "end point" is the best obtainable index of the state of deterioration of the fabrics. Some have expressed the opinion that the weighting of small failures is excessive. And with the individual assessment of minor damage. For example, four 1/4" holes in close proximity receive a score of 20, but one hole of equal combined area, or 1/2" diameter, receives only a score of 9. Decreases in wear score could also occur when two small holes combine to form a larger diameter hole, were it not for the policy of retaining the maximum score attained in such cases. The choice of an "end point" at which most of the garments are unserviceable, however, tends to minimize these effects. Steps taken to correct for biases resulting from the varying difficulty of detecting wear areas on different types of fabrics have already been mentioned.

Before a study is initiated to verify or revise the present weighting system, a review of procedures for setting up scoring systems is necessary. The empirical method already used is a tedious process and requires a large number of expert judges which are difficult to obtain. Thus the problem is to select a procedure which is most efficient for both reliability and ease of handling.

ON THE RELATION BETWEEN THE ENGINEER AND THE STATISTICIAN*

Joseph Mandelson
Chemical Corps Materiel Command

Increasing complexity of research, development and production problems demands new approaches, new tools, new methods of attack, lest ever-mounting complications slow or strangle scientific progress. These difficulties, generated by increased sophistication in scientific development and by the accelerated pace demanded by technical, economic, and military competition, have caused the engineer** to appraise, with becoming modesty, his own powers in his chosen field.

Time was when knowledge was so limited that one man could, almost literally, "know it all." The grasp of a Newton or a da Vinci over and in advance of the physical sciences of their day cannot possibly be approached by any one man over the vastly larger scope of science today. Yet this is required to permit significant progress today. How can we do it? We no longer expect great advances in science from the unaided efforts of the individual. No, science today advances mainly through teamwork, frequently by nationwide coordination of many teams all operating in the same field of interest. The team concept is reflected with increasing frequency in modern research, development and production organizations. Teams comprise groups of scientists or technologists who specialize in the disciplines pertinent to the problem.

Increasingly we find that one or more of the team is a statistician. The statistician is not new to the field of engineering; Charles Darwin brought data evaluation problems to the mathematician, Galton, for resolution. But the statistician has not really been associated with engineering problems on a scale commensurate with his ability to contribute, though a gratifying start has been made in applying the work of such pioneers as "Student", Pearson, Fisher, and Shewhart.

At first, and to this day, the engineer appeared slow to recognize the value of expert statistical guidance. Some statisticians have regarded this fancied weakness with some vehemence and self-righteous indignation, but if technology has not used statistics to its full potential the fault is neither the engineer's nor the statistician's alone. Perhaps most of it is due to human frailty.

---

* This paper was originally published in the May 1957 issue of Industrial Quality Control. Permission to reproduce it here is greatly appreciated by the editors.

** As used in this paper, the terms "engineering" and "engineer" refer broadly to the physical sciences, related application technologies, and practitioners in these fields. It is quite possible that relationships between statisticians and experts in subject matter fields other than "engineering" could profitable parallel lines suggested herein.

The problem is how to alert more of the scientific fraternity to the necessity for the team approach - in this case, the need for association of engineer and statistician. The statistician has a contribution to make in technology; it is up to him to "advertise" and "sell" it. In this, the statistician has been somewhat of a failure. In general, the statistician has not bridged the semantic and technical gap between himself and the engineer. Statisticians who published texts intended to popularize or spread the use of statistics among engineers, have been derided privately for writing "cook books" pandering to the low tastes of the statistically illiterate.. Such texts frequently open with a promise that only a modicum of mathematical background is required and then belie the statement in the next few pages, in line with the familiar usage in mathematical texts wherein an extremely complicated expression descends to a completely unlike formulation through the phrase "From which it is easy to see that:", or more simply "Hence:". Where engineers turned statisticians and published papers in technical journals, intending to popularize and illustrate the use of statistics by engineers, the publications all too frequently left the reader cold or frustrated. Perhaps such papers should first pass a critique by disinterested and uninformed engineers similar to the legendary stupid officers (no offense intended) who make apologetic appearance on the staffs of all famous military leaders as critical proving grounds for well-written battle commands.

Even when the need for cooperation between statistician and engineer is established, long-lived difficulties can be generated by the manner in which the relation between the two develops. Statisticians tend to overlook the fact that, while he may not have been particularly efficient at it, the engineer has actually formulated experiments, accumulated and evaluated data since his first day as a student. It is not surprising, therefore, that he feels reasonably competent in these fields. He views with complacency the generally satisfactory progress which science and technology have made with apparently little help from the statistician. But, though he has done quite well without statistics up to now, our engineer wants to be perfectly fair, broad-minded, and forward-looking. He does recall hearing about long-hair statisticians who can set up tests and evaluate the results better than engineers. Well, maybe....., in very complicated cases. "Tell you what; if I run into a really tough one where I can't find the best way, I'll call that statistician for advice --- what did you say his name was?"

As the situation developed, the engineer came to regard the statistician as a consultant. Now, the title may be one of dignity but the usage is occasionally less gratifying. Nromally a consultant's advice is accepted and used by the individual seeking these services. Occasionally, however, procedures recommended by a statistician are modified or even ignored by the engineer. There is little point in berating the engineer who refuses to accept the statistician's suggestions; this difficulty is relatively trivial. A more important problem by far is generated when the organization is such that the initiative as to the need for statistical consultation rests with the engineer. In other words, the statistician is consulted only when the engineer considers it necessary. Where this occurs, it is a grave weakness; the decision is left to the man least competent to make it. Until the statistician earns the engineer's full confidence, the engineer tends to regard the statistician somewhat as he would a child prodigy - interesting and clever in his way but not to be completely trusted with any real problem. So it is not surprising that the engineer frequently balks at, rejects or modifies recommended statistical procedures. In the case of a test design this may lead to failure of the test, mutual recriminations and distrust.

In this uncomfortable situation the statistician may take refuge in certain dodges, which evade the issues and tend to entrench him more firmly in the "ivory tower" which the engineer considers to be the statistician's normal habitat. Some statisticians insist that they will refuse to operate upon data generated in tests they did not design, nor will they stir if a test design they specified is modified in any way. It is hard to believe that such an adamant stand is actually maintained in practice but it is loudly advocated. Of course, in adopting this position, the statistician may lose golden opportunities to point out the havoc wrought in testing programs by poor test designs and then follow up with constructive recommmendations. Other statisticians suggest that the consultant confine himself solely to statistical aspects of the problem and avoid any discussion or speculation in the (non-statistical) technical subject matter phase. It is equally dubious that in any practical situation attitude is maintained. Certainly, the statistician should refrain from pretending to authority in a field not his own, but this should not prevent him from making suggestions as to the possible engineering meaning of his findings. There can be no objection to the statistician venturing into the technical field. However, final decision as to the engineering significance of statistical findings must be made by the engineer, except that such decision must not controvert the data as statistically evaluated. Equally there is no harm in the engineer offering comments to the statistician, particularly as regards the technical and economic realism of his statistical designs, provided it is understood that final decision in this case rests with the statistician.

The statistician must be sensitive to criticism which indicates that his test design is too expensive, complicated or time-consuming. He must be sure that the work projected in his testdesign is not more than is required to achieve the objectives set by the engineer. Many statisticians do not realize that running a factor once at each of two levels almost always involves much more work than running the factor twice at the same level. They count only the number of determinations to be made, not the operational changes for each determination.

The statistician should not accept in wordless resignation unilateral changes by the engineer in statistical test designs or data evaluation. Neither should the engineer surrender his birthright in fields (such as sampling and quality control) which are statistical in theory but engineeering in application. The statistician must not sulk in his tower as Achilles in his tent, nor should the engineer be silent when given a mathematical pattern devoid of engineering sense. On the contrary, given sufficient cause , let each scream to the heavens, for from honest controversy truth and understanding may emerge. The plain fact is that close cooperation is essential between the statistician and the engineer. Each must strive to explain to the other his needs or findings in the detail required for complete understanding by both of the problem, data, conclusions, and recommendations involved. The engineer has too often refused to make the mental effort required to comprehend the service the statistician can provide. The statistician has too often considered the engineer's problem an opportunigy to make a display of erudition rather than a contribution, confounding himself as much as he does the engineer.

Many of the faults properly laid at the statistician's door result from his ignorance of or lack of interest in the engineering subject matter field involved. At the same time some of the blame must be accepted by the engineer who, when faced by some statistician's brainstorm, fails to demand a clear explanation to demonstrate the value or utility of the procedure recommended by the statistician. This weakness merely encourages the statistician to commit additional sins against the engineer.

The need of the statistician to understand the engineering subject matter with which he will deal can best be appreciated through actual experience. Statisticians who are most competent in the application of statistics to technology almost invariably possess considerable educational accomplishment and experience directly in the engineering field. Very rarely do we find a top-notch statistician active in the engineering field who has had no previous formal engineering training. In this regard, the technological statistician is very much like a patent lawyer who is most successful when he holds an engineering degree, usually acquired prior to his legal training. It is not held that a statistician, as such, cannot eventually become very useful in the engineering field; it is simply much more difficult for the statistician to turn engineer than it is for then engineer to turn statistician.

There are many examples from actual practice wherein statistician and engineer contribute to the solution of pressing technical problems through cooperative effort in a manner which can best be described as interpenetration of the sciences involved. A single example will suffice to show clearly how interplay of intelligences operating in both fields, frequently in fine disregard of purist attitudes commonly adopted, can illuminate problem areas only dimly realized and effect important advances in the technology concerned.

The case in point started during a "coffee break" when the statistician heard a chance remark made by a production engineer that a certain arsenal found it easy to manufacture Widgit X* to prescribed quality requirements during the winter, but it was difficult to meet these requirements during the summer. This interested the statistician; it is strange that most of the production engineers knew about this but, after some trivial, half-hearted attempts at investigation, they gave it up as a bad job.

In this instance, however, the statistician provoked further discussions, later broadened to include their respective superiors. It was decided to investigate the truth of this assertion and, if true, to try to discover why we had trouble during the summer but not during the winter. At the time, the Widgit was no longer being manufactured, so that only information which could be gleaned from existing inspection and production engineering reports would be available as grist for the statistician's mill. Were the statistician a purist, he might have begged off on the plea that the data were not generated in accordance with a statistical test design, that the body of data might be incomplete in significant areas, that it was too voluminous, etc. In short, he had every excuse to refuse the job.

---

* The identity of Widgit X, its components, and the actual data to which we refer are not revealed for military reasons. However, it can be said that Widgit X was of utmost importance during World War II.

Instead, he welcomed the opportunity. His study immediately revealed a definite relation between average quality at the time of acceptance test and the date of manufacture. Further, he discovered that almost all malfunctions occurring during test were caused by failure of a pellet pressed from a mixture of two simple chemicals.

At this stage of the investigation, the statistician could do no more. He discussed his findings with the engineer, pointing out that it was obviously not the date which influenced Widgit quality but it was something associated with the date which, for technical reasons, would have an effect upon the Widgit. Two possibilities were offered by the engineer: temperature and humidity. The statistician explained to the engineer that, if temperature and humidity were related to date, both factors would undoubtedly be found to correlate statistically with Widgit quality. Since it had already been shown that date was related to quality, it would obviously follow that whatever characteristic related to date the engineer chose, it would also be found correlated with quality. (In making these statements, the statistician had begun to enter the field of engineering to explain to the engineer the statistical consequences of the engineer's decision.)

For chemical reasons, the engineer decided that the significant characteristic was probably relative humidity. In reply to questions put by the statistician, the engineer allowed that he expected the correlation between relative humidity and functionability to be high and inverse, rather than direct and low. No direct information was available as to the relative humidity at the Widgit assembly line; but by making suitable assumptions, it was possible for the engineer to draw a chart allowing calculation of humidity at the assembly points from temperature and outdoors ambient relative humidity.

At this point purists would have washed their statistical hands of the matter. No one pretended that this chart represented data which were generated as part of a statistical test design. There was some question as to the full extent of its validity; no way existed to answer such question with authority. Nevertheless, based upon this chart, a correlation was made between calculated relative humidity and Widgit quality expressed as percent effective. Fortunatley, what we lacked in precision of data was more than made up for by volume. As expected, relative humidity was found to be correlated inversely with quality with very high statistcal significance but the coefficient of correlation was disappointingly low, approximately -0.3. Were it not for the fact that the sample size ranged in the thousands, this correlation might never have become evident.

The statistician reported his findings to the engineer. Again he reminded the engineer that this relation between relative humidity and quality merely reflected the relation between date and quality; further, the actual correlation found, while highly significant statistically, was very much lower than the engineer had predicted. Neither the engineer nor the statistician could explain this. The engineer felt this might possibly have been caused by lack of truly precise humidity data; the statistician felt that it might have been due to the fact that the correlation was calculated as linear while it might have actually been curvilinear. It is interesting that each sought the explanation of the discrepancy in his own field - the engineer concerned himself with the humidity problem, the statistician with the question of linearity. Little could be done about

the former, but when the statistician ran a test for linearity he made the key discovery.

The correlation study had been based on grouped data. In examining the data for linearity, the statistician discovered that the negative correlation between quality and humidity was extremely high at low humidities. It became worse, though still significant, as relative humidity increased, but, for some strange reason, above 55% RH there was no significant correlation between quality and humidity. Now it was easy to understand why the overall correlation, which encompassed all classes of humidity, was lower than the engineer's expectations. But to explain why the correlation ceased above 55% RH was beyond the statistician.

At this point, knowing these facts, it would be interesting for each to ask himself, were he the statistician, what he would do. We have found quality to be related to date and, thereby, inversely related to relative humidity. The relationship was excellent at low humidities, grew worse at higher humidites and was completely lost above 55% RH. We knew that practically all malfunctions were due to failure of a pressed pellet of a mixture of two simple chemicals. These findings seemed to have very little practical significance. Those of us who are engineers might well question what further steps could be taken.

It is at this point that the truly competent statistician must rise to the occasion. He must use his knowledge of the engineering subject matter, however limited, to furnish guesses, wild guesses if need be, to catalyze the engineer's thoughts and help him determine why the correlation ceases above 55% RH. Above all, it was critical that the statistician recognize intuitively that the answer to this question was very probably the engineering crux of the problem. Certainly no ordinary engineer would, on his own, be interested in what appeared to be purely a statistical freak, let alone divine that it had any engineering importance.

In this case the statistician hazarded the guess that one or both of the chemicals, or perhaps an impurity therein, was characterized by some significant property immediately associated with the figure 55% RH, such that ambient humidities below 55% RH had one effect on pellet functioning while humidities above that figure had a different, probably opposite, effect. The question was: what was the characteristic property and what chemical material was involved? This question was the spark that ignited the engineer's intelligence. The physical characteristic involved was plainly the equilibrium relative humidity and the difficulty was pinned on a certain chemical impurity in one of the chemical constituents of the pellet. Upon searching the literature, it was found that the equilibrium relative humidity of this impurity at room temperature 54%. Now the phenomenon was completely explainable chemically, physically and statistically.

It became perfectly clear that to eliminate the variation in production quality caused by humidity, one of three courses would have to be taken: use chemical constituents free of impurity, or use the impure constituent but control humidity at the assembly points and within the Widgit container, or develop another component impervious to moisture. We didn't stop here, These findings were preliminary to even more important studies and findings with respect to the remaining stock of Widgits not exhausted by use in war. Joint studies by statistician and engineer enabled

us to estimate with utmost precision the useful serviceable life of the Widgits remaining in storage. Considering what we discovered in this and succeeding studies in connection with the Widgits in storage, it is not too much to say that they comprised the most important single group of statistical and mathematical studies carried on within our technical service in the past ten years. We learned how to manufacture better Widgits and what to expect with respect to their storage life. We were also able to develop a basic theory of lotting to insure homogeneity in production. It gave us the insight to develop the theory and practice of grand lotting in Chemical Corps surveillance. But perhaps most important of all, it laid the foundation for a proper relation between the statistician and the engineer at Chemical Corps Materiel Command.

Since that time, it has been understood that the top echelon of our Quality Assurance engineers require a good working knowledge of technological statistics and our top echelon of statisticians must have engineering degrees. Other professional personnel in the Quality Assurance Directorate, who may have little or no training in one of the two fields, are motivated to strengthen themselves in their weak area, for this is the road to advancement. At the same time, they are kept in closest contact with their counterparts in the opposite field so that all operate in accordance with the doctrine that each man makes final decision in connection with problems in his own area and offers such comments and suggestions to the other area as he may deem helpful. In any important matter subordinates bring their findings to their superiors who can speak with almost equal authority in either subject matter field.

The important organizational factor, probably unique in character, is that this relation of engineer and statistician is enforced. The engineer is not left to decide for himself whether he need consult with the statistician on data evaluation, test design, sampling, quality prediction, and the like. Nor is the statistician permitted to publish engineering conclusions in his studies unless these are acceptable to the responsible engineer.

The organization requires engineer and statistician to work together as coequal partners in the solution of quality assurance problems. Since both the engineering and the statistical groups are charged with responsibilities which insure enforced continual contact in most problem areas, one may well wonder whether this might not conduce to jurisdictional disputes. We can only reply that, in over ten years of operation, no such dispute has ever arisen. Though, in any given field of interest such as sampling, both engineering and statistical considerations are seemingly inextricably intertwined, the principles of the organization form the thin sharp line of demarcation: namely, every man makes final decision in his own professional field and has the right and is encouraged to offer suggestions, comments, and make all the mistakes he wants in the other man's field and with impunity, since responsibility for final decision as to acceptability of these comments rests with his team-mate in the other professional field. It works. Try it.

COMMENTS ON THE PAPER BY JOSEPH MANDELSON

A. Bulfinch
Picatinny Arsenal

"The paper 'On the Relation Between the Engineer and the Statistician' by Joseph Mandelson of the Army Chemical Corps is the best paper on this subject that I have heard. It reflects a great deal of thought on preparation and presentation. However, I have two questions: 'How can we have top executives in an established organization trained in statistics?' This of course is ideal and should be the organization's objective. 'But just how is this to be accomplished within the tenure of office of our present executives?' Surely we are not going to change executives because they are not trained in statistics. Neither should we add to their burdens by asking them to take a course in statistics. It should be enough that the executive assures himself that his first line supervisors are trained in statistics. This would imply that the supervisors stand ready to advise the executive on statistical matters and that the supervisors are themselves applying statistics and assisting their people to do likewise."

DESIGN OF AN EXPERIMENT IN THE RELIABILITY ANALYSIS OF A COMPLEX COMPONENT

James W. Mitchell
Frankford Arsenal

There is little that is new or novel in the mathematical aspects of reliability. However, the subject is receiving much attention in the fields of electronics, missiles and aircraft because of unique aspects of application and interpretation which arise in each new problem in these vital military areas. This paper treats with one of these unique areas of reliability estimation, namely, that of a relatively costly mechano-explosive device in the final stages of development. These devices are characterized by being self-destructive in use; hence they can be operated but once during their life.

Since the performance of the device cannot be tested before it is placed in service, final acceptance is usually based on quanitative performance tests on a small sample of a lot. Yet the device is too complex and costly to permit extensive reliability testing on the complete device sufficient to give even a fair estimate of over-all reliability. Examples of this type of device are emergency electrical or mechanical power sources and escape systems for aircraft, power sources and gas generators for guided missiles. They are characterized by an initiating mechanism, either electrical or mechanical; then an explosive train of primer, booster and propellant or explosive; and finally the energy output mechanism, a piston, gas turbine, dynamo, expanding bellows, etc. A device of this kind is expected to be ready for use when needed and yet to remain installed in some vital military equipment for months or even years until put to use. The high value of the life or equipment that the device powers or protects demands high reliability; certainly not less than 0.9999 or one failure in 10,000, better if attainable.

It should be interesting to examine the reliability estimate that could be based on the satisfactory performance of 100 units during the final evaluation of a new design. If this sample were the only source of reliability data and if none malfunctioned, binomial probability methods can be applied giving the following estimates.

The probability is 0.9 that R is not less than 0.977 or the probability is 0.995 that R is not less than 0.948.

Even a sample of 500 or 1000 units tested without failure would not suffice to guarantee a reliability of 0.9999. In fact a sample of something like 25,000 would be required.

If the failure of this type of device were caused only by environmental conditions and not by the presence of defects at the time manufacture, a sensitivity or increased severity test could be used to measure reliability. The test would employ the major environmental effect for the stimulus—if it were known. The trouble with this approach is the "ifs" which in most cases are not valid assumptions.

It should be apparent by now that more data are needed if a better reliability estimate is to be made. The only source of these data, outside of greatly increased testing, is background experience and knowledge on

similar subcomponents as used in the design requiring evaluation. Most devices of the kind under discussion are an assembly of series related subcomponents. These subcomponents can often be tested separately and often are a part which is used with but small variation in many different designs. There is therefore past experience on which to draw, or an opportunity to collect data on its performance. It then remains to complete the reliability model.

At this point an example will serve to illustrate one approach to a more complete reliability estimate. The example is based on an aircraft escape catapult. The following series of subcomponents can be identified in this item; sear and spring driven firing pin, primer, black powder booster, propellant charge and telescoped tubes to transmit the energy to the airman's seat. With this series relationship of subcomponents, a product relation can be assumed leading to the following reliability equation.

$$R = Q_1 \cdot Q_2 \cdot Q_3 \cdot Q_4 \cdot Q_5 \cdot Q_6$$

where

R = overall reliability
Q_i = reliability of the ith subcomponent

The difficulty with this equation is that the Q_1 are not simple single valued factors, but rather are functions of manufacture, input energy, temperature and environment, age, etc. All of these factors together will determine the subcomponent or the cataput reliability. All of them would have to be taken into consideration in a complete reliability model of a device during its service life. Although worth while and probably attainable with considerable effort, such a complete model is certainly beyond the scope of this paper. If the environmental and ageing effects are omitted, one is essentially considering the reliability of the item at the time of manufacture. A model for this condition will be suggested.

A number of variables remain which determine the value of Q for any subcomponent. These variables can be expressed in terms of their probabilistic effect on the subcomponent as follows:

1. The probability of a critical defect in all of the parts and the assembly of the subcomponent; for example, defective metal, missing parts or incorrect assembly.

2. The distribution of energies, expressed in some suitable form of the output of the preceeding subcomponent.

3. The conditional probability of failure of the subcomponent due to its variable sensitivity to ignition or actuation by the output of the proceeding subcomponent.

4. The probability of a critical defect during final assembly of the whole component.

It should be evident that variables 2 and 3 above result in an interaction which produces a single contribution to the probability of failure of the subcomponent. This interaction effect may be negligible for some subcomponents but quite significant for others. An important example for the

kind of devices considered in this paper is the interaction between the variable energy (velocity) of the firing pin and the variable sensitivity of the primer. More will be said of this later on. The fourth variable above will enter into the overall reliability of the device but once.

The net effect of the above variables is that the Q_1 can be expressed as a function of two factors, one representing the incidence of critical defects and the other the interaction of the subcomponents. This is expressed as follows:

$$Q_i = (1-p_i)\ (1-p_{i/i-1})$$

where

p_i = Probability of a critical defect in the lot of subcomponent i.

$p_{i/i-1}$ = Conditional probability of failure or sensitivity of the i th component to the variable energy output of subcomponent i-1.

Now if $q = 1 - p$, $Q_i = q_i q_{i/i-1}$, and the overall reliability equation can be expressed as follows:

$$R = q_1 \cdot q_2 q_{2/1} \cdot q_3 q_{3/2} \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \; q_i q_{i/i-1} \circ q_a$$

The term q_a represents the final assemply reliability.

Equations are of no value without data to apply them to. However, no simple set of rules can be suggested to obtain the estimates of subcomponent reliability for insertion into the above equation. An example seems best at this point to illustrate possible approaches to the problem. It is hoped that this will be sufficiently suggestive to enable others to use this method. Subcomponent reliability estimates for a type of cataput device as described earlier, are given in the following table. Also indicated are the sources of these data.

Subcomponent Reliability q_i

Subcomponent	q_i	Source of Information
Sear	.99996	About 20,000 of different designs tested without failure due to this component under normal loading. All of nearly similar design and tolerance.
Firing Pin	.9986	No failures in 500 tested of a new light pin design.
Primer	.99999+	Major defect ratio known from millions made
Booster	1	Laboratory tests with reduced charges show that limit causing failure to ignite far below inspection limits. Also inspection nearly foolproof for low charges.
Propellant	1	

(Continued)

Subcomponent	q_i	Source of Information
Telesc. Tubes	.99996	
Fabrication	.99996	Same as for the Sear above.

Subcomponent Interaction $q_{i/i-1}$

Subcomponent	$q_{i/i-1}$	
Sear	—	
Firing Pin	—	No interaction since firing pin is powered externally. Sear is only a release.
Primer	.99995	From the overlap of the distributions of firing pin energy and primer sensitivity
Booster	1	Laboratory tests mentioned above with primer, booster and propellant show no significant interaction within normal loading limits.
Propellant	1	
Telesc. Tubes	1	No interaction if required minimum propellant charge is present.

Overall Reliability (from previous equation) R = .9984

A few remarks are required on the above table. A reliability value of 1 is used to indicate a very high degree of reliability estimated from the large margin of safety uncovered in laboratory tests. The reliability values for q_i, except for the primer, were estimated from an extension of the inverse solution of the incomplete Beta-function ratio $0.5 = I_o\ (c, n-c+1)$, published in the book "An Engineers' Manual of Statistical Methods" by L. E. Simon, John Wiley and Sons, 1941.

It will be apparent that the above means of obtaining data for the reliability equation depends heavily on the judgment of the engineer. It will also be evident where further development is necessary to provide a design with a reliability approaching 0.9999. In the case of the primer-firing pin interaction, the given value was obtained from the known statistics of the primer sensitivity and the distribution of firing pin energies as measured on a sample by the copper indent method. A design improvement was made by the elimination of a light metal sealing disc over the primer. This resulted on about 40 per cent increase in primer sensitivity and reduced the probable incidence of misfires in this device from 0.00005 to 0.000001 or less. The overall reliability is obviously determined largely by the lack of data on the firing pin system. This can be rectified as more test data is obtained with the new pin or by a carefully planned laboratory program on this part.

The contribution of this reliability model in terms of new information may appear trivial. Its real contributions are two in number. First, with the collection of sufficient test data on the many subcomponents of designs

in use, the model will enable a design engineer to think in terms of reliability as the design and testing phases of a new model proceeds. When new concepts are incorporated or there is a low reliability figure on certain subcomponents, the model should serve as a red flag. It should serve to indicate the need for more trusted parts or for a parallel study of the part to measure and improve its reliability. Second, when a new design is completed and its output performance tested and found to be satisfactory, there will also be a good basis on which to offer an estimate of reliability that might be expected when the item goes into production. However, like all statistical answers, it will be only an estimate, not a guarantee. Yet it should be better than any other estimate that could be made on a developmental component.

PUNCHED CARD COMPUTING OF F-TESTS

G. H. Andrews, J. Dominitz, G. T. Eccles, C. J. Maloney, and C. W. Rigg
Army Chemical Corps

INTRODUCTION. For several years the computation of analysis of variance by punched card methods has been performed routinely by personnel of these laboratories Standard Sperry Rand equipment is used, consisting of a UNIVAC 120 electronic computer, tabulator, multi-control reproducing punches, sorters, and key punches. A brief description of the computation procedure employed was included in a paper (1) presented by Dr. Clifford J. Maloney at the first of these conferences in Octobe 1955. In that paper it was reported that research was underway to "devise methods of determining observed F values by calculation on the UNIVAC 120 so that the choice of appropriate error terms could be based on any selection of pooling rules". The present paper summarizes the results of these efforts to date.

BACKGROUND. It is well recognized that many of the computations involved in the analysis of variance are more efficiently performed on punched card equipment than on desk calculators. Transformation of the original data when appropriate, summing over the various treatment combinations to form all the desired tables, squaring, summing of squares, "correction" of sums of squares, and division to get mean squares are all obvious applications of the tabulator or UNIVAC 120 computer. An application not so obvious is the calculation of the variance (F) ratios and their associated probabilities on the 120.

The calculation of variance ratios is in itself a very simple operation, given the mean squares for the several sources of variation and a designated error term. Not so simple is the choice of error when the various proposals for pooling are considered. More will be said about this aspect of the problem later.

Calculation of the probability associated with a given variance ratio and its accompanying degrees of freedom was believed to be desirable if more complete mechanization were to be sought, but it was realized that this would be a difficult achievement on a computer with the limited storage and program capacity of the UNIV 120. In view of the limited number of mathematicians in our laboratories and an ever-present backlog of research assignments for them, this problem was referred to the Statistical Laboratory of Iowa State College, Ames, Iowa, under the terms of a contract between that institution and the Chemical Corps. At Ames, this problem received the attention of Dr. H. O. Hartley, who devised several limited solutions (2). These solutions are limited in this respect. The first method presented, while comparatively easy to program, is restricted to even degrees of freedom for both numerator and denominator of F. The second method, while somewhat more difficult to program, is less restricted in scope since only the denomina degrees of freedom are required to be even. Since many analyses involve a comparison of only two treatments (one d.f.), the second method was chosen as being the more practical.

---

(1) Maloney, Clifford J. "Punched Card Computing of Analyses of Variance." Proceedings of the First Conference on the Design of Experiments in Army Research, Development and Testing, Office of Ordnance Research Report No. 57-1, June 1957, pp. 97-127

(2) Hartley, H. O. "Programs for Computation of Incomplete Beta Function and F-Integral on Machines with Limited Storage." Technical Report No.12, Statistical Laboratory, Iowa State College, Ames, Iowa, April 10, 1956.

PROCEDURE. Under present operating conditions, two computer runs are required to calculate variance ratios and their associated probabilities, following the computation of the mean squares. At the mean square stage of computation, each of the sources of variation is represented by one card which containes the "corrected" sum of squares, mean square, degrees of freedom and some word or symbol identifying the source of variation. Prior to the first computer run, the cards containing the mean squares are sorted so that the card containing the error mean square will enter the computer first. This mean square and its degrees of freedom (rounded to the next lower even number, where necessary) are then punched from storage into each of the cards containing effects to be tested. On the same card pass, the ratio of the two variances is computed and punched into each card. If this ratio is less than 1, codes representing the letters "NS" (abbreviation for non-significant) are punched in a appropriate position in the card. The second computer run gives the actual probabilities associated with the F ratios greater than 1 and their respective pairs of degrees of freedom.

At this point a few words about the choice of error may be appropriate. In certain analyses, choice of error (or errors) is not difficult, there being only one variance suitable for error and no opportunity for pooling. These analyses present no problem. Other analyses, due to the nature of the factors involved or the extent of breakdown of the analysis offer pooling opportunities. At the present time our procedure in these cases is limited to testing all effects against the highest order interaction. When subsequent pooling is not suggested by this preliminary test, no further steps are necessary. If pooling of certain variances is advised by the preliminary test, the pooling and subsequent tests of significance are then accomplished on the desk calculator.

After the second computer run, the cards containing the sums of squares and probabilities for each source of variation are then resorted into their proper order, listed, and given to the statistician for his examination and interpretation. The final listing for a typical analysis is shown in Table I.*

SIGNIFICANCE. Since the procedure outlines above is a fairly recent development, advantages have yet to be fully realized. They are expected to be two-fold: (1) computation of the F ratios and associated probabilities by punched card methods will lighten the load of the statistician by carrying the mechanical processing several steps beyond that previously attempted; and (2) completed analyses of variance similar to that shown in Table I can be typed in somewhat altered but final form on the card-operated flexowriter, thereby saving the time of the typist in the preparation of the final report.

The sole disadvantage is the requirement that the degrees of freedom for the denominator variance be even. The practice of rounding downward to the next even number results in little or no change in the computed probability when ample degrees of freedom are included in the denominator mean square. When only a few degrees of freedom are represented by the denominator mean square, manual reference to the computed tables of F is advisable.

MATHEMATICS. It has been mentioned that the major share of the mathematical analysis of this problem was undertaken by Dr. H. O. Hartley in his paper (3). Those who are familiar with this reference will recall that the success of the program depends upon the calculation of the incomplete Beta Function, since

$$PR\left\{F \leq F_c\right\} = I_x(a,b) = \frac{\Gamma(a+b)}{\Gamma(a)\,\Gamma(b)} \int_0^x u^{a-1}(1-u)^{b-1}du$$

* Table I is at the end of this article.

(3) Hartley, op. cit.

WHERE $a = \frac{1}{2}\nu_2$, $b = \frac{1}{2}\nu_1$, AND $x = \frac{a}{a+b\,F_o}$.

Alternate methods of computing the incomplete Beta Function are advanced. The first approach, somewhat more limited of the two, is based upon the fundamental recursion formula for $I_x\,(a,b)$ which is

$$I_x\,(j,i) = x\,I_x\,(j-1,i) + (1-x)\,I_x\,(j,i-1)$$

$$\text{where } I_x\,(1,i) = 1 - (1-x)^i$$

$$\text{and } I_x\,(j,1) = x^j.$$

The second computational scheme suggested, and the one actually pursued by the Statistics Branch, is based on a different recursion formula

$$I_x(j,i) = I_x(j+1,i-1) + \binom{i+j-1}{j} x^j (1-x)^{i-1}$$

The summation of this formula leads to two representations of the incomplete Beta Function according as ν_1 is or is not even. If ν_1 is even, summation yields

$$I_x(a,b) = \sum_a^{a+b-1} \binom{a+b-1}{j} x^j (1-x)^{a+b-1-j}$$

If ν_1 is odd, the result takes the form

$$I_x(a,b) = \sum_a^{a+b-3/2} \frac{\Gamma(a+b)}{\Gamma(j+1)\Gamma(a+b-j)} x^j (1-x)^{a+b-1-j} + I_x(a+b-1/2,1/2)$$

No precise statement can be made concerning the computational time because it is so dependent upon the number of degrees of freedom. We have run some combinations of degrees of freedom taking nearly 5 minutes to compute, but the norma run averages between 3 and 4 second, certainly a great improvement upon searching through a table and resorting to interpolation.

TABLE I

EFFECT		DF	MS 1	F	PROB	MS 2	SOS		
REPLICATION	R	1	44731	9104	.00415	4913	44731	1	46
VARIETY	V	2	9514	1936	15578	4913	19029	2	46
AGE	A	3	26039	5300	319	4913	78117	3	46
TREATMENT	T	3	7057	1436	23507	4913	21173	4	46
V X A		6	6951	1414	22946	4913	41706	5	46
V X T		6	4629	942	NS	4913	27774	6	46
A X T		9	8153	1659	12676	4913	73382	7	46
V X A X T		18	4772	971	NS	4913	85896	8	46
ERROR		47	4913				230919	99	
TOTAL		95	6555				822727	99	

LIFE TESTING *

Benjamin Epstein
Wayne State and Stanford Universities

I. INTRODUCTION. In recent years there has been an increasing interest in developing statistical and probabilistic methods which can be used to improve the design and analysis of life and fatugue tests. Such tests are destructive, time consuming, and expensive and there is a great need for statistical methodology, which will enable the experimenter to squeeze the maximum amount of information from whatever data are accumulated.

A characteristic feature of life and fatigue test data is that information becomes available in an ordered way. Thus, if we place n items on test, we can discontinue testing long before all n items have failed. In particular, we may decide to stop testing as soon as we have a preassigned number ($r \leq n$) of failures; or we may decide to stop testing by a preassigned time T_0; or according to a suitable sequential rule. In which, by taking advantage of the time ordered nature of the data, enable the experimenter to reach a decision in a shorter time or with fewer observations, than would be possible otherwise.

II. THE EXPONENTIAL DESTRIBUTION AND ITS ROLE IN LIFE TESTING. In this paper we limit ourselves about exclusively to considering problems of life testing under the assumption that the life X is described by a probability density function of the form

$$f(x;\theta) = \frac{1}{\theta} \exp(-x/\theta), \quad x > 0, \quad \theta > 0. \tag{1}$$

In (1), x is life measured in appropriate units (for example, hours) and $\theta = E(X)$ is the mean life expressed in appropriate units. There is evidence that the lives of electron tubes or the time intervals between successive breakdowns of electronic systems, are to a first approximation, random variables having the density (1).**

The beauty of the assumption of the exponential distribution of life is that it makes it possible to apply the theory of Poisson processes. Furthermore one can by almost trivial changes generalize all the results to the case where the conditional rate of failure is some function of time, $Z(t)$, rather than a constant as in the exponential case. The theory thus extended has validity over a wide area including most cases of practical interest. One should bear this in mind as one reads the rest of this paper in which we describe some of the results that have been obtained in the exponential case.

---

* Preparation of this paper was supported in part of the Office of Naval Research and the Office of Ordnance Research.

** Recently we have prepared a manuscript entitled "The exponential distribution and its role in life testing." In it we have considered various stochastic failure models and the life distributions associated with these models. From these considerations one sees that the exponential distribution and some other closely related families of distributions must play a fundamental role in life testing.

A few references relevant to the exponential assumption are:

(1) D. J. Davis, "An Analysis of Some Failure Data," Journal of the American Statistical Association 47, 113-150, 1952.

(2) C. R. M Tuttle and A. R. Frank, "Inventory Control Methods Applied to Electronic Tubes," a paper included in Logistics Papers, Issue 8, November 16, 1951-February 15, 1952. Appendix I. Issued by the George Washington Logistics Project.

(3) E. S. Rich, "Experience with Receiving Type Vacuum Tubes in the Whirlwind Computer Project," Report R-194, Project Whirlwind, Servomechanisms Laboratory, Massachusetts Institute of Technology, February, 1951.

(4) Aeronautical Radio Incorporated, "Investigation of Electronic Equipment Reliability as Affected by Electron Tubes," Inter-base Report No. 1, March 15, 1955.

(5) D. R. Cox and W. L. Smith, "On the Superposition of Renewal Processes," Biometrika 41, 91-99, 1954.

III. SUMMARY OF RESULTS ON TESTING HYPOTHESES IN THE EXPONENTIAL CASE. (One Sample Situation). Although time limitations make it impossible to give details, it does seem appropriate to summarize what is known in the exponential case. Results are given in a variety of situations including cases in which items on test may or may not be replaced, where the life test may be truncated either after a fixed number of failures have occurred, or after a fixed amount of time has elapsed, or where the decision is made on a sequential basis. In all situations one finds ideas and methods involving Poisson processes to be very useful.

Suppose that we wish to test $\theta = \theta_o$ against $\theta = \theta_1$ $(\theta_o > \theta_1)$ with prescribed

Type I error (producer's risk) α and prescribed Type II error (consumer's risk) β then tables, formulae, and general theory are given principally in 4 papers:

(1) B. Epstein and M. Sobel, "Life Testing," Journal of the American Statistical Association 48, 486-502, 1953;

(2) B. Epstein, "Truncated Life Tests in the Exponential Case," Annals of Mathematical Statistics 25, 555-564, 1954;

(3) B. Epstein and M. Sobel, "Sequential Life Tests in the Exponential Case," Annals of Mathematical Statistics 26, 82-93, 1955;

(4) B. Epstein, "Statistical Problems in Life Testing," Proceedings of the Seventh Annual Convention of the American Society for Quality Control, 385-398, 1953.

In paper (1) the principal results were as follows: The "best" estimate based on the first r out of n failures is given by

(2) $$\hat{\theta}_{r,n} = \left[\sum_{i=1}^{r} x_i + (n-r)\, x_r\right] / r$$

in the non-replacement case and by

$$\hat{\theta}_{r,n} = nx_r / r \tag{3}$$

in the replacement case. n is the number of items initially placed on test and r is the number of failures. The p.d.f. of $\hat{\theta}_{r,n}$ is in either case given by

$$f_r(y) = \frac{1}{(r-1)\,!} (r / \theta)^r y^{r-1} e^{-ry/\theta}, \quad y > 0 \tag{4}$$

$$= 0 \text{ , elsewhere}$$

and further it is easily shown that $W = 2r\hat{\theta}_{r,n}/\theta$ is a random variable which is distributed as chi-square with 2r degrees of freedom (χ^2 (2r)).

The expected waiting time for the r'th failure is given by

$$E(X_{r,n}) = \theta \sum_{j=1}^{r} 1/(n - j + 1) \tag{5}$$

in the non-replacement case and by

$$E(X_{r,n}) = r\,\theta/n \tag{6}$$

in the replacement case.
A test procedure having size (Type I error; producer's risk) equal to α is described by a region of acceptance of the form

$$\hat{\theta}_{r,n} > C = \theta_o \chi^2_{1-\alpha} (2r) /2r \quad . \tag{7}$$

It should be noted that the p.d.f.'s of $\hat{\theta}_{r,n}$ and of $W = 2r\hat{\theta}_{r,n}/\hat{\theta}$ are independe of n and that the acceptance region described by (7) is also independent of n. Thi means that, in the exponential case, all tests and estimates based on the first r out of n failures (n arbitrary) are equally good and the only choice among procedu depends on the relative cost of n (the number of items on test) and $E(X_{r,n})$, the ex pected waiting time for the r'th failure in a sample of size n.

Formula (7) gives for each r a test procedure for which the probability of accepting a lot having mean life θ_o is given by $L(\theta_o) = 1 - \alpha$. If one wishes the O.C. curve for the procedure to be such that the probability of accepting a lot hav ing mean life θ_1 is given by $L(\theta_1) \leq \beta$, then r (and hence C in (7)) must be chos appropriately. Details for doing this and proofs of all the results given above may be found in reference (1) .

Much more detailed discussion, proofs, and tables can be found in the ONR report, "Some Tests Based on the First r Ordered Observations Drawn from an Exponential Population," by B. Epstein and M. Sobel, Stanford University Technical Report No. 6 (N6onr-25126) and Wayne University Technical Report No. 1 (Nonr-451(00)).

In paper (2) the interest is focussed on finding truncated test procedures in either the replacement or non-replacement case. With n items placed on test it is decided in advance that the life test will be terminated at $\min(X_{r_o,n}; T_o$ where $X_{r_o,n}$ is the time at which the r_o'th failure occurs and T_o is the truncation time beyond which the experiment will not be run. Both r_o and T_o are preassigned in advance. If the experiment is terminated at $X_{r_o,n}$ (i.e., r_o failures occur before T_o) then the action taken is to reject. If the experiment is terminated at time T_o (that is, the r'th failure occurs after time T_o) then the action taken in the language of hypothesis testing is to accept. These test procedures are characterized by three functions: $E_\theta(r)$, the expected number of failures before reaching a decision; $E_\theta(T)$, the expected waiting time to reach a decision; and $L(\theta)$: the probability of accepting if θ is the true value of the mean life. Formulae for these three functions and relevant theoretical considerations are given in paper (2). It is further shown in Section 3 of paper (2) that the test procedure obtained for the non-replacement case in paper (1) can be viewed as a truncated test (in the sense that it need not necessarily run until the first r failures are observed) and that the test procedure obtained for the replacement case in paper (1) is precisely the same as that obtained in paper (2).

Useful formulae and tables are given which enable the experimenter to determine the appropriate truncated life test meeting the following conditions:

(i) the O.C. curve is to be such that $L(\theta_o) \geq 1 - \alpha$ and $L(\theta_1) \leq \beta$ (where $\theta_o > \theta_1$) and

(ii) the life test must be terminated by time T_o at the latest.

The formulae and tables enable one to determine the two integers R_o and n, where r_o is in the usual language of sampling inspection the rejection number and n is the sample size (that is number of items placed on life test).

It is appropriate to mention at this point that the truncated test procedures just described are good rules of action in cases where the underlying distribution of life is not necessarily exponential. More precisely, we mean the following: Suppose that an acceptable lot of electron tubes is one for which the probability of failing before some time T_o is $\leq p_o$ and that an unacceptable lot is one for which the probability of failure before some time T_o is $\geq p_1$ $(p_1 > p_o)$ and suppose we want the O.C. curve to be such that $L(p_o) \geq 1 - \alpha$ and $L(p_1) \leq \beta$. It is then an

easy matter to find a sample size n and rejection number r_o such that we will accep the hypothesis that $p = p_o$ if the number of defectives (failures before T_o) in the sample $\leq (r_o - 1)$ and reject the hypothesis that $p = p_o$ (accept $p = p_1$) if the number of defectives in the sample $\geq r_o$. This test procedure clearly is truncated and has the property that $L(p_o) \geq 1 - \alpha$ for any distribution $F_o(x)$ which is such that $\int_0^{T_o} dF_o(x) \leq p_o$ and $L(p_1) \leq \beta$ for any distribution $F_1(x)$ which is such that $\int_0^{T_o} dF_1(x) \geq p_1$. If, in particular, $F_o(x) = 1 - e^{-x/\theta_o}$, with

$$\theta_o = T_o / \log\left(\frac{1}{1 - p_o}\right) \text{ and } F_1(x) = 1 - e^{-x/\theta_1} \text{ with } \theta_1 = T_o \log\left(\frac{1}{1 - p_1}\right),$$

the test procedure just described has the property that $L(\theta_o) \geq 1 - \alpha$ and $L(\theta_1) \leq \beta$. Recalling that the rule of action can be written as <u>accept</u> if $\min(X_{r_o,n} ; T_o) = T_o$ and <u>reject</u> if $\min(X_{r_o,n} ; T_o) = X_{r_o,n}$, we have precisely the truncated procedure which one gets in the exponential case when testing θ_o against θ_1 with $L(\theta_o) \geq 1 - \alpha$ and $L(\theta_1) \leq \beta$. But from the foregoing argument the test procedure is the appropri one to use when we wish to distinguish between two distributions $F_o(x)$ and $F_1(x)$ with

$$\int_0^{T_o} dF_o(x) \leq p_o = 1 - e^{-T_o/\theta_o} \text{ and } \int_0^{T_o} dF_1(x) \geq p_1 = 1 - e^{-T_o/\theta_1}.$$

For all such cases $L(F_o) \geq 1 - \alpha$ and $L(F_1) \leq \beta$.

Sequential life tests in either the replacement or nonreplacement cases are given for testing $\theta = \theta_o$ against $\theta = \theta_1$ $(\theta_o \geq \theta_1)$ with Type I errors $= \alpha$ and Type II error $= \beta$. A continuous analogue of the sequential probability ratio test of A. Wald can be used. In paper (3) we give formulae for the O.C. curve; for the expected number of failures, $E_\theta(r)$; and for the expected waiting time, $E_\theta(t)$, before a decision is reached. We also give a table of values of $E_\theta(r)$ for certain choices of θ_o/θ_1, α, and β.

In paper (4) a summary is given of results in papers (1), (2), and (3) and an example is worked out comparing various procedures for testing θ_o against θ_1 with prescribed α, β.

IV. PROBLEMS OF ESTIMATION. Useful results on estimation are given in the following paper and reports:

(1) B. Epstein and M. Sobel, Wayne University Technical Report #1, written under ONR Contract Nonr-451(00), March 1, 1952.

(2) B. Epstein, Wayne University Technical Report #2, written under ONR Contract Nonr-451(00), July 1, 1952.

(3) B. Epstein and M. Sobel, "Life Testing," Journal of the American Statistical Association 48, 486-502, 1953.

(4) B. Epstein and M. Sobel, "Some Theorems Relevant to Life Testing from an Exponential Distribution," Annals of Mathematical Statistics 25, 373-381, 1954.

(5) B. Epstein, "Life Test Estimation Procedures," Technical Report No. 2, written under OOR Contract DA-20-018-ORD-13272, July, 1954.

(6) B. Epstein, "Simple Estimators of the Parameters of Exponential Distributions when Samples are Censored," Annals of Statistical Mathematics 8, 15-26, 1956.

(7) A. E. Sarhan and B. G. Greenberg, "Tables for Best Linear Estimates by Order Statistics of the Parameters of Exponential Distributions from Singly and Doubly Censored Samples," Journal of the American Statistical Association 52, 58-87, 1957.

It was mentioned in the part of our discussion devoted to testing hypotheses that "best" point estimates of the unknown parameter θ are given in references (1) and (3) by $\hat{\theta}_{r,n}$ (See formulae 2 and 3). In references (1) and (5) we also give confidence intervals and tables useful in computing confidence intervals from knowledge of $\hat{\theta}_{r,n}$ (which is based on knowing the times of occurrence of the first r failures when n items are placed on test at time $t = 0$). In references (2) and (5) it is shown that in the exponential case, approximate estimation procedures of very high efficiency can be given even if one disregards the first $(r-1)$ failure times $x_1, x_2, \ldots x_{r-1}$ and retains only x_r, the time of occurrence of the r'th failure. In fact, in the replacement situation, the estimate based on x_r coincides with the one based on knowing $x_1, x_2, \ldots x_r$, since x_r is a sufficient statistic.

Frequently life test data become available from several experiments. It is shown how to combine all of this information into an optimum point or interval estimate in references (4) and (5). Additivity of the χ^2 distribution plays a fundamental role in these considerations. An additional problem considered in detail in reference (4) is the simultaneous estimation of the parameters A and θ in the two-parameter exponential distribution described by the p.d.f.

$$(8)\quad f(x;\,\theta,\,A) = \begin{cases} \frac{1}{\theta} e^{-(x-A)/\theta} & ,\ \text{for } x > A \\ 0 & ,\ \text{otherwise}. \end{cases}$$

Specifically, a formula is given for a uniformly minimum variance unbiased point estimate of A, and a procedure is given for finding a confidence interval for A which is optimal in a certain sense. This confidence interval can also be used as a test of significance for A. A distribution such as (8) with $A > 0$ would correspond to a situation where every item lives for at least a length of time A.

In reference (5) we also discuss the following: estimation when the life test data are truncated in time; estimates having a prescribed precision; a number of approximate estimation procedures; estimates of quantity, etc..

In reference (6) we extend results in references (2) and (4). Reference (7) gives further results from a somewhat different point of view. Some of the result in (6) and (7) are closely related to each other.

V. TWO SAMPLE TESTS. Up to this point we have considered problems of testing or of estimation where one has a single sample. It frequently happens that one has two samples from each of two populations and wishes to find out on the basis of the samples whether or not the populations are essentially different. Two papers dealing with questions of this kind are:

(1) B. Epstein and C. K. Tsao, "Some Tests Based on Ordered Observations from Exponential Populations," Annals of Mathematical Statistics 24, 458-466, 1953;

(2) B. Epstein, "A Sequential Two Sample Life Test," Journal of the Franklin Institute 260, 25-29, 1955.

In reference (1), we have the following non-sequential life test situation: Let $x_{11} \leq x_{12} \leq \ldots \leq x_{1n_1}$ and $x_{21} \leq x_{22} \leq \ldots \leq x_{2n_2}$, be two random samples (S_{n_1} and S_{n_2}) from populations having p.d.f.'s $f(x;\,A_1,\,\theta_1)$ and $f(x;\,A_2,\theta_2)$ respecitvely, where $f(x;\,A,\,\theta)$ is the two-parameter exponential described by (8). Let S_{r_1} and S_{r_2} be the sets of the r_1 and r_2 smallest observations of S_{n_1} and S_{n_2} respectively, then tests are given for the following situations:

(a) H_1: To test $\theta_1 - \theta_2$

(assuming A_1 and A_2 are known).

(b) H_2: To test $\theta_1 = \theta_2$
(assuming $A_1 = A_2$, but that the common value is unknown).

(c) H_3: To test $\theta_1 = \theta_2$.

(d) H_4: To test $A_1 = A_2$.
(assuming θ_1 and θ_2 are known).

(e) H_5: To test $A_1 = A_2$.
(assuming $\theta_1 = \theta_2$, but that the common value is unknown).

(f) H_6: To test $A_1 = A_2$

(g) H_7: To test $A_1 = A_2$ and $\theta_1 = \theta_2$.

The kinds of hypothese tested remind one of analogous problems for the normal distribution where μ plays the roles of A and σ plays the role of θ. Thus H_5 is the analogue of the classical problem of Student in the two-sample case and H_6 is the analogue of the Behrens-Fisher problem.

In reference (2) we consider a sequential two sample life test of the following kind:

Suppose that a user of electron tubes is given two lots of tubes and that he wishes to choose the lot having the greater mean life on the basis of a life test made on sample of tubes drawn from each lot. To make the problem precise we assume that tubes in lots 1 and 2 have a life time described by the p.d.f. (1), with associated mean lives θ_i, i = 1, 2. Generally speaking, $\theta_1 = \theta_2$, or $\theta_1 > \theta_2$ or $\theta_1 < \theta_2$. If $\theta_1 = \theta_2$, we are indifferent as to the ranking assigned to the two lots. If $\theta_1 > \theta_2$, we should prefer to have the decision precedure lead to the (correct) assertion that lot 1 has the greater mean life. Similarly if $\theta_2 < \theta_1$, we should prefer to have the decision procedure lead to the (correct) assertion that lot 2 has the greater mean life.

A measure of how lot 1 compares with lot 2 is given by the ratio $u = \max(\theta_1, \theta_2)/\min(\theta_1, \theta_2)$. In most practical problems, the experimenter would like to have a high probability of properly ranking the two lots if u equals some specified number u_o (>1). The sequential procedure given in reference (2) has the following properties:

(i) when u = 1, the probability of calling $\theta_1 > \theta_2$ (or $\theta_2 > \theta_1$) is equal to .5;

(ii) as u increases, the probability of assigning the proper ranking to the two lots increases;

(iii) when $u \geq u_o$, the preassigned ratio, then the probability of ranking the two lots incorrectly is less than or equal to some preassigned small α.

The sequential procedure and formulae for the probability of assigning a wrong ranking and for the expected number of items failed in the course of reaching a decision follow directly from the paper by M. A. Gershick, "Contributuions to Sequential Analysis I," Annals of Mathematical Statistics 17, 123-143, 1946. It is also interesting to note that another way of finding the basic formulae of this paper is to rephrase the decision problem in the terminology of the problem of the ruin of the gambler. For details on the latter problem one should refer to W. Feller, "An Introduction to Probability Theory and Its Applications," New York; John Wiley and Sons, Inc., 1950, pp. 282-288.

Two recent papers which are relevant to the subject discussed in this part of our paper are:

M. Sobel, "Statistical Techniques for Reducing Experiment Time in Reliability Studies," Bell System Technical Journal 35, 179-202, 1956 and M. Sobel and M. J. Huyett, "Selecting the Best One of Several Binomial Populations, "Bell System Technical Journal 36, 537-576, 1957.

VI. NON-PARAMETRIC LIFE TEST RESULTS. Up to this point we have in the main discussed life test results obtained under the assumption that the underlying p.d.f. is exponential. In closing, we wish to mention some recent results of a non-parametric nature in the two-sample case. The papers are:

(1) C. K. Tsao, "An Extension of Massey's Distribution of the Maximum Deviation Between Two-Sample Cumulative Step Functions," Annals of Mathematical Statistics 25,587-592, 1954.

(2) B. Epstein, "Tables for the Distribution of the Number of Exceedances," Annals of Mathematical Statistics 25, 762-768, 1954.

(3) B. Epstein, "Comparison of Some Non-Parametric Tests Against Normal Alternatives to Life Testing," Journal of the American Statistical Association 50, 894-900, 1955.

(4) M. Sobel, "On a Generalization of Wilcoxon's Test With Applications to Reliability and Life Testing," a paper presented at the NYU-RCA Conference on Reliability, April 17-19, 1957.

In these four papers the basic hypothesis under test is that the two c.d.f.'s $F(x)$ and $G(x)$ are such that $F(x) = G(x)$. There exist in the literature several procedures for doing this. Three of these tests are: The maximum deviation test of the kind generally associated with the names of Kolmogorov and Smirnov and also recently worked on by F. M. Massey, Z. W. Birnbaum and other others; the exceedance test developed by Gumbel and von Schelling; and the rank-sum test developed by Wilsoxon. In life testing we are interested in tests of this kind where n_1 items from one c.d.f. and n_2 items from the second c.d.f. are placed on life test and where the life test is terminated as soon as a preassigned number of failures $r_1 (\leq n_1)$ and $r_2 (\leq n_2)$ occur in the two populations respectively. The titles of the papers are indicative of which particular truncated non-parametric life tests were under consideration.

VII. CONCLUSION. Our purpose, in this lecture, has been primarily to describe research, the bulk of which has been done under contracts in Life Testing with the Office of Naval Research, and the Office of Ordnance Research. We have emphasized primarily the case where the life test distribution is exponential, since this has proved to be a natural starting point for research in this field. We have not attempted to make an exhaustive survey of all of the work done in the field of life testing because we felt that this lies outside the scope of our present discussion. * We feel that a good beginning has been made in this important new area of research. Although the results are being used primarily by people working with physical and electronic problems, we are sure that they can also be used effectively in many other fields. In particular we hope that some of you may find the ideas and methods useful in your own work.

---

* This is being done in a Handbook on Statistical Methods in Life Testing, currently being prepared by the author under an ONR contract.

CHANGES IN THE OUTLOOK OF STATISTICS BROUGHT ABOUT BY MODERN COMPUTERS

H. O. Hartley
Iowa State College

1. INTRODUCTION. The topic of this talk may be regarded by some to imply something that is undesirable.

Computers are, after all, mechanical tools. In spite of what we see advertised in glowing colors about these 'Electronic Brains' they cannot do anything on their own account and the human brain has to provide all their thinking. The idea then that this mechanical slave should influence the thoughts of scientists may appear altogether dangerous and undesirable. After some reflection we may be inclined to admit that this new tool may have an influence on our methods of computation. However, computational methods are regarded by many statisticians as a sort of second class area, a necessary evil to obtain numerical answers, a trivial m tter not worthy of discussion!

Characteristic of this attitude is a casual remark of a British statistician:- He was apparently being bored when a group of fellow statisticians were discussing convenient computational procedures for doing regression work, and remarked that 'the computational procedure most convenient to him was to proceed to the computing room and tell them to get on with it!'

It is true that computers are mechanical tools. As such they certainly affect the computational aspects of statistical analysis. However, will they reach the heart and soul of statistical methodology, will they influence the statistical outlook? I think they will.

Let us look at an analogy. Soon after Roentgen discovered X-Rays their medical potentialities were realized. They became a mechanical tool to help physicians in their diagnosis, but did they influence the outlook on medical treatment? No doctor would dispute today that they did:- Quite apart from the fact that X-Ray therapy is a recognized treatment, the mere fact that X-Rays make it easier to diagnose internal troubles has influenced the clinical outlook fundamentally. The mere fact, then, that high speed computers can carry out computations much faster influences our evaluation of statistical methods. In my brief report on this influence I will confine myself to two aspects:-

I. The influence of computers on the evaluation of mathematical functions required in statistics.

II. Their influence on the analysis of empirical data and data processing.

2. THE EVALUATION OF MATHEMATICAL FUNCTIONS REQUIRED IN STATISTICS. Mathematical functions are used in statistical analysis in numerous ways. We may distinguish two main uses, however:-

1. The use of tables of the important statistical distribution functions in the form of tables to draw inferences from statistical summaries of data.

2. The use of elementary and higher mathematical functions for the purpose of variate transformation in the course of statistical analysis.

In both of these situations the evaluation of functions is greatly facilitated by the use of electronic computors. It will be seen that in the first situation we are speaking of tables of statistical functions. The effect of high speed computers on the role of tables in numerical analysis has recently come up for frequent discussion. In fact it has been suggested that tables of mathematical functions will not be required in the high speed computations of the future. Two reasons are proposed:-

a. On practically all computors subroutines for high speed internal computation of most mathematical functions are available.

b. Tables of functions, if internally stored in the computor, will occupy an unacceptably large proportion of the high speed access storage. On the other hand, if tables are currently passed through the machine by the input media,(that is, tapes or cards) the scanning for the required value is unacceptably slow except in special circumstances.

At a recent meeting at the M. I. T. these questions concerning the role of tables were discussed. It was felt that the above view is justified but only for computations on high speed computors. On the other hand it was felt that printed tables of functions would still be required for a long time to come. Some of the reasons given were:-

c. That tables would be required in pilot computations of a research nature and indeed in the planning of the larger scale computations of the high speed computors.

d. That for some time, at any rate, immediate access to the services of high speed computors would not be universal.

As far as statistical usage is concerned all these arguments, a, b, c and d, clearly apply to tables of variate transformations and similar functions summarized under item 2.

Variate transformations are applied during data processing and if the data are analyzed on a high speed computor the function can be computed 'ad hoc' in the machine by a subroutine. Tables of these functions will, however, be required for pilot computations on desk-computors.

The situation concerning statistical distribution functions (item 1) is, however, different. Normally such tables are consulted by the statistician after the data have been analyzed and after statistical summaries have been computed. For example, tables of percentage points of F are used to guide the research worker when he is drawing inferences from his analysis of variance mean squares. So the statistician requires the F tables, and normally not the machine. If someone should suggest that the computing machine should take over from the statistician the task of drawing scientific inferences, I think we would all agree that this would be virtually impossible as well as undesirable. In general, therefore, tables of the distribution functions are used after the data analysis. However, we must not overlook that they must sometines be evaluated during data processing. The weighting tables in probit analysis are an example and so are other iterative procedures in maximum likelihood estimation as well as in other statistical procedures.

In assessing the effect of high speed computors on the evaluation of mathematical functions we must, therefore, bear in mind two points:-

1*. They facilitate the tabulation of statistical functions for printed circulation.

2*. They facilitate the ad hoc computation of statistical and other mathematical functions required during data processing.

The principles concerning 2* are well known to numerical analysts and are not characteristic of statistical functions. We therefore confine ourselves here to the case 1*, i.e. to the preparation of tables, and we are here mainly concerned with tables of statistical distribution functions. In Schedule 1 we have classified such tables by type of function (1, 2, 3 and 4) and method of computation (A, B and C). Let us consider the effect of high speed computors on the computation of these tables and illustrate it with examples:-

Type 1. Where a convenient mathematical formula is available for a more or less straightforward computation good tables are usually in existence. Examples are the tables of the standard statistical functions, the normal or t-distribution, χ^2 and F-tables, Binomial and Poisson distributions and certain derived functions based on these. These tables have been computed (usually without the help of high speed computors) from their exact mathematical formulas. Occasionally effective approximations by standard distributions have been invented which make a tabulation unnecessary. Witness Fisher's Z-transformation which does away with the need for tables for the normal correlation coefficient r (David 1937).

SCHEDULE 1.

Computation of Statistical Tables Classified by Type of Mathematical Function and Method of Computation

	Method of Computation A	B	C
Type of function	Tabulation from exact formulas	Approximate formulas tested by mathematical analysis and/or trail tabulation reduce function to standard or special tables.	Approximate formulas tested by Monte Carlo reduce function to standard or special tables.
1. Convenient formulas for the tabulation are available	All standard statistical tables, e.g. normal, t, χ^2, F etc.	Approximations by standard statistical tables are used to save tabulations e.g. Fisher's z-transformation of the correlation r	Not required
2. The function can be tabulated at considerable computational labor	Distributions with involved formulas e.g. Order statistics, mean deviation, measures and ratios of normal dispersion, mean square consecutive difference	Approximate evaluation from fitted expansions (Gram Charlier, Edgeworth, etc.) Pearson type curves Reduction to normal by variate transformation e.g. Distributions of $\sqrt{b_1}$ and b_2 in normal samples	Not required but frequently practiced by classical Biometrika School for purposes of illustration
3. The function is multiparametric and its exact tabulation is impractical	Impractical e.g. Bartlett's test for heterogeneity of variance depends on k parameters	The approximations only depend on a few computable combinations of the parameters e.g. Normal, χ^2 and F approximations to multivariate test criteria Distribution of quadratic forms and ratios thereof	Randomization test criteria examined on sub-samples from complete combinatorial distribution e.g. Randomization tests in analysis of variance approximated by F-tests
4. The distribution problem is untractable by mathematical analysis	Not available	Not available	Graduation of Monte Carlo distributions e.g. Rank correlations for dependent rankings Queuing problems and simulation procedure

Type 2. We now turn to distributions which can only be computed with considerable effort, but once computed can be conveniently tabulated. Here the arithmetic high speed facilities of electronic computors are of great help. Examples of this kind are the distributions of order statistics and functions thereof (e.g. range) both for normal and other samples. Measures of normal dispersion thereof are exemplified in the table of the standardized extreme deviate $(x_n - \bar{x})/\sigma$ in normal samples computed by Grubbs at the Ballistic Research Laboratories on the ENIAC shortly after World War II. This was one of the first statistical tables produced by a high speed computor. Earlier during World War II ordnance research required the computation of a particular criterion for trends in quality control of production. This was the ratio of the mean square consecutive difference to the sample variance:-

$$\delta^2 = \sum_{i=1}^{n-1} (x_{i+1} - x_i)^2 \Big/ \Sigma\, (x_i - \bar{x})^2$$

The mathematical theory of this distribution was developed by the late D. Von Neuman, R. H. Kent, H. R. Bellinson and B. I. Hart. The table was computed at Aberdeen Proving Ground. This piece of work is characterized by an ingenious combination of exact but complex formulas evaluated for small sample sizes n and Pearson-Type approximations which become more effective for larger sample sizes. The latter method of approximating to the distribution by such devices as Pearson type curves, Gram Charlier series and the like was frequently used in the past. Examples are the Pearson-Type approximations to the distributions of the normal moment ratios $\sqrt{b_1}$ and b_2, the mean deviation and similar criteria examined by the classical Biometrika School. It is clear that with the help of electronic computors we shall be able to replace more and more of these approximations by exact computation. Notice that these approximations are usually least satisfactory for small sample sizes, n, and that it is for small sample sizes that we can certainly expect a high speed computor to evaluate these distributions; for if more sophisticated methods* fail we can always integrate numerically over the n-dimensional sample space of the parental distribution.

Type 3. We now come to a very important class of distributions and one which makes tabulation a most difficult problem:- I am speaking of distributions which involve a large number of parameters and variates. An example is the multivariate normal distribution which, for k variates, involves ½ k(k-1) correlation coefficients. Usually there is no inherent difficulty in computing the integral for any specified set of parameters and limits of integration on the high speed computor but the multi-parameter multivariate tabulation is clearly impractical. (For example the 10-variate normal would involve 10 + 45 = 55 arguments.) Other examples are the multivariate likelihood criteria such as Bartlett's test for heterogeneity of variances which strictly speaking depends on the k degrees of freedom ν_i; i=1, 2...k, of the k sample

---

*Von Neumann's formulas mentioned above, e.g. Geary's 1947 recurrence computations (iterations) for $\sqrt{b_1}$ and b_2

mean squares. Similar criteria arise in multivariate analysis. In these situations effective approximations have been suggested which depend only on a few simple functions of the parameter. For example Bartlett's χ^2 - approximation to his test criterion depends on k, the number of mean squares and a somewhat improved approximation (Hartley 1944) on the two additional quantities

$$\Sigma \frac{1}{\nu_i} - \frac{1}{N} \quad \text{and} \quad \Sigma \frac{1}{\nu_i^3} - \frac{1}{N^3} \quad \text{where } N = \Sigma \nu_i$$

Similar χ^2 and F approximations for a large class of likelihood criteria were evolved by G. E. P. Box (1949).

The future role of high speed computors lies in checking the validity of these approximations by evaluating the exact distribution for a representative set of parameter combinations. Once such approximations have been established the criterion can usually be referred to standard tables. Sometimes special tables of the approximate function depending on fewer parameters may have to be computed. (See e.g. Biometrika Tables for Statisticians, Vol. I Table 32). The question arises: What are we to do if no such lower parametric approximations exist? Indeed this problem of tabulation arises with multivariate functions in general and is not confined to statistical functions. When this difficulty was discussed at the M.I.T. conference it was suggested that the function should be computed 'ad hoc' on the high speed computor for the particular combinations of parameters for which it is required in the course of the computations. Such a procedure may be suitable in certain problems of applied mathematics. However, at the present time a statistician would think twice before he lets the computor carry out (say) a 39 dimensional integration for the purpose of (say) testing the significance of a computed test criterion. Such procedures will have to wait considerable development of the present day computing techniques. In this category, therefore, statisticians will be forced to provide suitable approximations to their multi-parametric criteria in spite of the help that high speed computors are able to render.

Type 4. We finally come to distribution problems which are untractable by mathematical analysis. In such situations the stochastic process which generates the distribution is well defined but the probability distribution which is generated by it cannot be described by mathematical formulas which are sufficiently simple to permit the numerical evaluation of the distribution; put briefly we should say in such cases:- 'The statistician cannot do the problem.' Here the high speed computor comes to his help. 'Monte Carlo' - methods can often be used in such cases. These methods have been known for a long time but because of the tremendous computing effect which they entail they have not been seriously used for the solution of distribution problems until the advent of high speed computors. Let me explain in an over-simplified example what these methods entail:- Consider a random sample of n drawn from a normal population.

We know that the sum of squares of deviations of n observations x_i viz. $\Sigma(x_i-\bar{x})^2$ is distributed as χ^2 for n-1 degrees of freedom. Suppose we wanted to do this distribution problem by Monte Carlo. We would have to proceed in 3 steps:-

1. Generate inside the machine random samples each containing n values x_i from the normal $N(0,1)$.

2. For each of the samples compute the sum of squares of deviations $\chi^2 = \sum_{i=1}^{n} (x_i - \bar{x})^2$. Each sample of n values of x_i therefore only yields a single value of the statistic χ^2.

3. Make a grouped frequency distribution (histogram if you were to draw it) of the values of χ^2.

As more and more values of χ^2 are added to this distribution the grouped frequency distribution should approximate to the exact distribution

$$\frac{1}{\Gamma\frac{(n-1)}{2}} \exp\left[\frac{-\chi^2}{2}\right]\left(\frac{\chi^2}{2}\right)^{\frac{n-3}{2}} \cdot d\left(\frac{\chi^2}{2}\right)$$

Monte Carlo methods are therefore conceptually very simple:- They follow to the iota the very definition of the 'random s pling' distribution of a statistic (such as the above χ^2 statistic).

However, in practice, considerable improvements are necessary before this method can claim to produce a table of a distribution function to anything like an acceptable accuracy. As you all know very large sample sequences are needed:- This can be easily seen by a simple application of what is known as the Kolmogoro Smirnov criterion for goodness of fit. Suppose you have computed a cumulative distribution by Monte Carlo from N sample sequences. How close to your answer will be the true cumulative distribution. The above criterion tells you [See e.g. Massey F. (1951)] that with 99% confidence you can say that the maximum error in your computation is within $\pm 1.63/\sqrt{N}$, that is, the error decreases with $1/\sqrt{N}$. Conversely suppose you want your table to have just 3 accurate decimals with a 99% confidence, how many sequences do you require? The answer is immediately obtained from the equation $1.63/\sqrt{N} = 5 \times 10^{-4}$ or $N = 10^8 (1.63/5)^2 = 10.6$ million samples!!

It was soon realized that if Monte Carlo methods were to be used at all that new methods of reducing the sample sequence must be developed. Alternatively these methods have been termed methods of variance reduction. Considerably progress has been made on these lines as we may witness from the recent 'Symposium on Monte Carlo Methods' (Gainsville 1954). We may mention here at least 5 methods designed to achieve reduction in variance, namely:-

Importance of Correction Sampling	(H. Kahn and A. W. Marshall)(195
Multistage Sampling	(A. W. Marshall) (1954)
Conditional Sampling	(J. Tukey and Trotter) (1954)
Antithetic Variables	(J. M. Hammersley and K. W. Morto (1955)
Control Variables	(E. C. Cieller and H. O. Hartley (1954)

For details of these methods the reader is referred to the papers in question (see list of references). May it suffice here to say that with all these methods sampling can be reduced considerably at no loss of precision and that the relative merits of these methods depend on the circumstances of the sampling problem and on the gadgetary of the high speed computor which is available. It is of interest that most of the above methods are closely related to devices which sample surveyors use when sampling life populations, i.e., stratification, regression estemates, optimum allocations and the like. Designers of sample surveys have always been concerned with reducing the variance of estimates at constant cost or sample size. We may in the future look forward to further blending of eff ts between the sample surveyor and the mathematical 'Monte Carlist.'

We must not forget to mention here the considerable work which has been carried out on the other aspects of the Monte Carlo computations, notably the automatic generation of the random samples by the high speed computor:- As is well known, the starting point is usually the generation of random numbers or random digits. Numerous methods of generating these in the form of pseudo-random numbers are described or listed in the Gainesville Symposium. The next step is usually to obtain continuous uniform variates by composing random digits as the decimal digits of the uniform variates ranging between 0 and 1. These are then interpreted as probabilities and are in turn transformed to random variates x. If we are concerned with normal samples as in the above example we transform the random probability value u to a normal deviate with the help of the familiar normal ogive. Mathematically the relation is

$$u = (2\pi)^{-1/2} \int_{-\infty}^{x} e^{-(1/2)t^2} \, dt$$

This transformation is accomplished by loading into the machine a subroutine which computes the normal % point x which corresponds to a 'tail area' of u. When we wish to generate random samples from other distributions we require subroutines for computing their % points. In the future development of Monte Carlo methods therefore we shall require first computing techniques for % points for any probability level for all the parental distributions we wish to sample from. The library of such computing programs is fast increasing.

Let me conclude this section by giving you just a few details of a [illegible]tion problem. This problem is virtually intractable by mathematical [illegible] but at least an approximate solution to it could be found by Monte [illegible]putations:- I am speaking of the distribution of Spearman's Rank [illegible] for dependent rankings. Let me explain this concept in terms of [illegible]le. Suppose an expert judge is called upon to judge the comparative [illegible] 10 pieces of fabrics of the same kind. He is unable to attach [illegible] value of the wear to the pieces but he is able to place them in [illegible] merit by judging their relative wear, as shown in the second line [illegible] that follows.

	Fabric Piece										
	A	B	C	D	E	F	G	H	I	J	
Rank order by judge	6	9	2	1	4	3	5	10	8	7	$= u_i$
Rank order by objective measure	3	10	1	4	5	6	2	9	7	8	$= v_i$

The pieces can also be subjected to an (expensive) objective test and their wear ranked on this as shown in the third line of the table above. The question arises whether there is any correlation between the judges and the objective rankings. Spearman's rank correlation r_s is in fact the simple correlation coefficient between the pairs of rank numbers, i.e.

$$r_s = \left(\Sigma\, w_i v_i - \frac{n(n+1)^2}{4}\right) \Big/ \frac{n(n^2-1)}{12} = 1 - 6\Sigma d_i^2 \,/\, (n^3-n) \quad \text{where } d_i = u_i - v_i.$$

Its distribution is well known in the 'Null case', i.e. when it is assumed that there is no correlation. In the case where there <u>is</u> correlation, however, its distribution is virtually untractable. To tackle the problem by Monte Carlo, large numbers of dependent rankings* were generated on a computor and Monte Carlo distributions computed for various degrees of dependence. It was noticed that the variance and shapes of these distributions depended on the degree of correlation, a dilemma well known from the normal correlation coefficient. For the latter, Fisher solved all the difficulties by his ingenious z-transformation. So the same was tried for Spearman's correlation with similar results:- The z-transform, $z_s = \tan h^{-1} r_s$, were found to be approximately normally distributed with standard deviations approximately given by: $1.0296/\sqrt{n-3}$ independent of ρ. Their mean values, on the other hand, are functions of ρ but independent of n.

The table below shows the standard deviations of the Monte Carlo distributions of z_s for 833 samples of size n = 30 and 500 samples of n = 50. They are compared with the fitted compromise value $1.0296/\sqrt{n-3}$.

Standard deviations of the Monte Carlo distributions of z_s

$\rho =$	.1	.2	.3	.4	.5	.6	.7	.8	.9	$1.0296/\sqrt{n-3}$
n = 30	.191	.194	.201	.193	.202	.197	.195	.202	.216	.198
n = 50	.143	.155	.154	.150	.152	.157	.146	.151	.153	.150

---

*The n pairs of rankings u_i, v_i were generated as the rank-numbers in a sample of n pairs x y of normal variates with correlation coefficient ρ Monotonic distortions of the x and y scales which leave the distribution of r_s invariant make the results apply to a much wider class of parental distributions.

More detailed results will shortly be published in a paper by E. C. Fieller, H. O Hartley and E. S. Pearson in Biometrika, where they will be joined by similar results on Kendall's Rank Correlation and the correlation computed with the help of the Fisher scores (Fisher and Yates Tables, Table XX). With the help of these results we can compare two rank correlations (z_s) by an exact test of significance, combine two measures of rank correlation for more precise estimation, analyze sets of rank correlation by normal theory analysis of variance. These Monte Carlo computations were carried out some 5 years ago, partly on ordinary Hollerith Punched Card models and partly on the 'Ace' at the National Physical Laboratories in England, and you will note that with equipment which would be regarded as very modest today, very short sequences of samples were generated. In the future the more abundant resources of computing equipment and methods of variance reduction should result in Monte Carlo computations of higher precision. The above study shows, however, that appropriate variate transformations and simple graduations of Monte Carlo distributions can be most effective and it is these tools that are likely to be used in future work of this kind.

3. ANALYSIS AND PROCESSING OF DATA. The second main activity of electronic computors is the statistical analysis of emprircal data. The problems arising here are quite different from those encountered in the computation of tables. It is convenient to distinguish two types of computational tasks.

1. Statistical 'Analysis of Data'

Under this category would fall such items as analysis of variance and covariance of experimental data, multiple regression and correlation analysis, probit analysis, etc.

2. Processing of mass data

Here we would be concerned with such activities as the tabulation of the results from sample surveys and Census tabulations. Borderline cases with industrial and commercial requirements such as inventory control also arise here.

We confine ourselves here to 1, i.e. to the statistical analysis of data and single out the analysis of variance* to examine the influence of high speed computors as well as the difficulties which arise in their use. It is fairly characteristic of the general trend.

The analysis of variance of data arising from and experiment is a numerical procedure which is very frequently performed in numerous centers. In spite of the considerable volume of computation expended on this activity most of the work is still carried out on desk computers, and this is even true of many centers at which the services of a high speed general purpose computer are available. The reason for this is undoubtedly the great

---

* The subsequent section is based on Hartley, H. O. (1956)

variety of experimental designs, each of which gives rise to a different type of analysis of variance each applied to a small body of data. There is no difficulty in setting up and testing suitable programs every time data from a new design are ready for analysis, but in so far as the time and effort of doing this is usually much greater than the effort of completing the analysis of variance on a desk computer, there is clearly no point in enlisting the high speed machines.*

It is obviously foolish for an expert to spend (say) 2 days writing and testing a brand new program for a particular design, then, for the machine to complete the analysis in (say) 2 minutes, whilst a competent desk computer could have completed the work in (say) 3 hours.

The efforts which are at present being made to overcome this dilemma of programming are well known:- They consist of the standardization of the analysis to simple basic operations. Already there are in use statistical interpretative routines and these incorporate subroutines for Analysis of Variance. These are programs which are set up and tested once and for all and can then be used for any new design with the addition of a small steering program. This basic analysis of variance calculus for the high speed computer differs from the familiar desk computer instructions. The latter are designed to save arithmetic, the former aim at standardizing the procedure to a few operations which are used in a simple logical sequence. Standardization is the key note even if this entails an increase in the arithmetic, after all the latter is of little concern to the high speed computor.

The basic analysis to which that of other designs may be reduced is that of a general factorial experiment.

For convenience we confine ourselves to a $k = 3$ factor experiment. Let x_{tij} denote the experimental result from t^{th} level of factor 'T', i^{th} level of factor 'I' and j^{th} level of factor 'J'. The symbols T, I and J will also denote the number of levels for each factor so that $t = 1, 2, \ldots, T$, $i = 1, 2, \ldots, I$ and $j = 1, 2, \ldots, J$. This complete analysis of variance of the T x I x J results into its $2^3 - 1 = 7$ components is shown in Table 1.

Table 1. Analysis of variance for 3-factor experiment

Component	Degrees of freedom
T	(T - 1)
I	(I - 1)
J	(J - 1)
T x I	(T - 1) (I - 1)
T x J	(T - 1) (J - 1)
I x J	(I - 1) (J - 1)
T x I x J	(T - 1) (I - 1) (J - 1)

* The comparative clerical labor of preparing the data for input into the respective machines, although by no means negligible, is not discussed here as this depends on the details of the organization of the computing center

For the corresponding sums of squares we shall require the familiar notation for group totals, viz.

$$X_{.ij} = \sum_{t=1}^{T} x_{tij} \qquad \text{and likewise } X_{t.j}\ X_{ti.}$$

$$X_{..j} = \sum_{t=1}^{T} \sum_{i=1}^{I} x_{tij} \qquad \text{and likewise } X_{.i.}\ X_{t..}.$$

The sums of squares can be obtained by repeated application of the following operators which will be explained in terms of factor T.

(1) Operator Σ_t $\equiv$ sum over all levels of $t = 1, 2, \ldots T$ whilst keeping the other subscripts constant

(2) Operator D_t $\equiv$ multiply all items by T and subtract the result of Σ_t from all items

Operator $(\ )^2$ $\equiv$ form the sum of the squares of the items inside the brackets and divide by the number of items.

For example if we apply the first two operators to the original set of results x_{tij} we have

$$\Sigma_t(x_{tij}) \equiv \Sigma_t\, x_{tij} = X_{.ij}$$ (the total for the i, j combination (of factors I and J

$$D_t(x_{tij}) \equiv T\, x_{tij} - X_{.ij}$$ (the deviate of x_{tij} from the i, j (mean multiplied by T.

The above simple operators represent the first two lines in the schedule of operations shown in Table 2 which gives complete formulas for the totals and deviates resulting from the sequence of operations $\Sigma_t D_t \Sigma_i D_i \Sigma_j D_j$ applied to the data x_{tij}. The seven sets of deviates finally reached in lines 9 to 15 are finally subjected to the Mean Square Operation $(\ )^2$ and the results are the 'Sums of Squares of Deviations' (all multiplied by TIJ) for the seven components of variance shown in the fifth column of Table 2. Table 3 (below) illustrates these operations with the help of a simple example in which $T = 3$, $I = 3$ and $J = 2$. It is these figures, i.e. the number of levels in

Table 2. Schedule of operations for three factor analysis of variance

Line	Operator	Applied to values in lines	Will form totals or deviates	Deviates used for analysis of variance components
1	Input		x_{tij}	
2	Σ_t	1	$X_{.ij}$	
3	D_t	1 and 2	$Tx_{tij} - X_{.ij}$	
4	Σ_i	2	$X_{..j}$	
5	Σ_i	3	$TX_{t.j} - X_{..j}$	
6	D_i	2 and 4	$IX_{.ij} - X_{..j}$	
7	D_i	3 and 5	$TIx_{tij} - IX_{.ij} - TX_{t.j} + X_{..j}$	
8	Σ_j	4	X	
9	Σ_j	5	$TX_{t..} - X$	T
10	Σ_j	6	$IX_{.i.} - X$	I
11	Σ_j	7	$TIX_{ti.} - IX_{.i.} - TX_{t..} + X$	T X I
12	D_j	4 and 8	$JX_{..j} - X$	J
13	D_j	5 and 9	$TJX_{t.j} - JX_{..j} - TX_{t..} + X$	T X J
14	D_j	6 and 10	$IJX_{.ij} - JX_{..j} - IX_{.i.} + X$	I X J
15	D_j	7 and 11	$TIJx_{tij} - IJX_{.ij} - TJX_{t.j} + JX_{..j} - TIX_{ti.} + IX_{.i.} + TX_{t..} - X$	T X I X J

Table 3a. Numerical example of a 3-factor experiment with $T = 3$, $I = 3$, $J = 2$ and sequence of operators $\Sigma_t\, D_t\, \Sigma_i\, D_i\, \Sigma_j\, D_j$

		j=1			j=2		
		i=1	i=2	i=3	i=1	i=2	i=3
x_{tij}	t=1	4	2	3	6	2	4
	2	5	6	2	3	1	2
	3	2	4	3	4	2	5
Σ_t		11	12	8	13	5	11
D_t	t=1	1	-6	1	5	1	1
	2	4	6	-2	-4	-2	-5
	3	-5	0	1	-1	1	4
$D_i\Sigma_t$		2	5	-7	10	-14	4
D_iD_t	t=1	7	-14	7	8	-4	-4
	2	4	10	-14	-1	5	-4
	3	-11	4	7	-7	-1	8
$D_jD_i\Sigma_t$		-8	19	-11	8	-19	11
$D_jD_iD_t$	t=1	-1	-10	11	1	10	-11
	2	5	5	-10	-5	-5	10
	3	-4	5	-1	4	-5	1

		j=1	j=2
$\Sigma_i\Sigma_t$		31	29
Σ_iD_t	t=1	-4	7
	2	8	-11
	3	-4	4
$D_j\Sigma_i\Sigma_t$		2	-2
$D_j\Sigma_iD_t$	t=1	-11	11
	2	19	-19
	3	-8	8

$\Sigma_j\Sigma_i\Sigma_t$		60		
$\Sigma_j\Sigma_iD_t$	t=1	3		
	2	-3		
	3	0		
		i=1	i=2	i=3
$\Sigma_jD_i\Sigma_t$		12	-9	-3
$\Sigma_jD_iD_t$	t=1	15	-18	3
	2	3	15	-18
	3	-18	3	15

each factor, that may vary from experiment to experiment and need to be conveyed to the machine.

Table 3b. Analysis of Variance of Data, Table 3a

Components	TIJ (S.o. squares)
T	6
I	78
J	4
T x I	186
T x J	182
I x J	182
T x I x J	46

The analysis of variance of many other designs can be reduced to the above factorial analysis by simple steering programs. This is shown by Hartley, H. C (1956).

It is clear, from the table, that at the end of the operation we shall have a complete record of all the residuals in the machine and this is quite contrary to desk computor practice except in the special case of a 2^k factorial. Here the residuals are identical with the 'contrasts' for main effects and interactions which are usually computed in a 2^k factorial experiment.

We now come to the question as to what influence will this operational calculus of analysis of variance have on the statistical procedures. I think the fact that the residuals are all easily and automatically computed will make it possible to

(a) Examine individual residuals for unusual features which would otherwise get lost in the sums of squares.
(b) Compound from the residuals individual contrasts of particular interest (e.g. contrasts between special treatment combinations).
(c) Compare the residuals obtained with different variate transformations with regard to additivity.

In a recent paper in Biometrics, Sir Ronald Fisher (1954) has pointed out the importance of the 'additiveness of the transformed variate when various controllable, or uncontrollable factors, the effects of which are to be analyzed, are varied'. Fisher's paper is concerned with the analysis of variance of quantal data, but his point clearly applies to variate transformations in general. Since the study of the additivity resulting from various variate transformations is considerably facilitated by the high speed computor we venture to predict that in the future data will be analyzed in a dual manner as follows:-

(a) Data from a particular experiment would be subjected to an analysis

of variance in the above manner using that variate transformation which, on the information to date, is the most appropriate.

(b) Simultaneously with the analysis in (a) the data would be subjected to alternative variate transformations suggested by alternative theories and the study of their residuals would suggest metameters resulting in better additivity and hence improve the current knowledge of theoretical background of the data.

The first analysis would provide quantitative estimates of the amount of variation attributable to the various factors under consideration. The second analysis would provide information on how to improve the metameters in which such variation should be measured in future experiments.

It may be asked why the analysis of variance of (a) should not be carried out with that variate transformation which yielded the best additivity in (b). Such procedures require special caution as they may lead to biased estimation. They certainly bias the estimation of error if that variate transformation is selected for which the error residual sums of squares is a minimum.

A few words may be added on other instances of data analysis.

In multiple regression analysis, for example, similar trends may be expected to occur. It is well known that the high speed computor is of considerable help here, particularly in multiple and non-linear regression analysis. The machine makes it possible to fit various regression laws suggested by alternate theories. For example, the yield from certain chemical reactions are governed by differential equations (usually assumed linear with simple exponential type of solutions) that could be fitted. It may now be proposed to modify the reaction theory by introducing additional terms allowing for effects previously regarded as negligible. Or, indeed, it may be proposed to fit yield surfaces obtained from a fundamentally different theory. Again I would suggest the dual analysis mentioned above, namely

(a) The current estimation of the coefficients in that regression law currently considered the most appropriate and their standard errors.

(b) An examination of the error residuals obtained from alternative laws suggested by new theories on the data.

It is clear that (b) must be confined to alternate laws with a theoretical justification as the residual sum of squares can obviously be made as small as possible by fitting a mathematical artifice to the data.

An important special case arises when the alternative regression laws all arise as 'special cases' of a 'general regression law'. For example, if we have the general regression law which is linear in the parameters

$$y = b_1x_1 + b_2x_2 + \cdots + b_kx_k$$

Now in this situation the question is sometimes posed whether certain of the independent variates are 'really necessary' and whether they should be 'discarded'. Every effort must, of course, be made to obtain all the

information that the theoretical background of the data can provide. However, there are situations when no such information can be obtained and the task is to get some indications from the data. Various test procedures for discarding independent variates have been tried in the past. Most of these depend on an a priori hierarchy of importance of the x_i. For example, if we try to determine a polynomial regression and are undecided what the degree of the polynomial should be, we may use for the x_i orthogonal polynomials and, starting from the highest degree x_k discard all 'insignificant x_i' until we reach the first 'significant' one. This procedure clearly depends on the hierarchy of the polynomial degree which in many situations is a sensible one, if only because it follows the logic of a Taylor expansion of a general analytic law. Where no such hierarchy of the x_i can be introduced the following procedure has been suggested:-

Fit first all the k independent variates $x_1 \dots x_k$.
Fit next all possible selections of k-1 variates out of the k.
Fit next all possible selections of k-2 variates out of the k ; and so on
Until last each of the x_i is fitted singly.
There will therefore be

$$1 + \binom{k}{k-1} + \binom{k}{k-2} + \dots + \binom{k}{1} = 2^k - 1$$

regression fits. For each of these compute the residual mean square s^2 and select the law for which s^2 is a minimum.

The following are the reasons why this procedure has not been used frequently in the past.

1. The computations are prohibitive.
2. There is some doubt as to whether the minimum residual mean square should be used as a criterion for selecting the law.
3. When the application of the criterion has selected a regression law all least squares estimates based on the same data are biased.

High speed computors remove objection 1. Indeed there are already programs in existance which will compute the 2^k-1 regression laws. However, we must remember that objections 2 and 3 still remain. Concerning objection 3, Kempthorne (1955) makes an interesting observation. Although he is concerned with other criteria he says 'perhaps the best that can be done, given a set of data, is to divide the two sets at random using one set to estimate the dependency relation and the other set to test the validity of this discovered relation'. This suggestion is similar to, although not identical with, the dual analysis suggestion made above. By splitting the issue of finding the 'best law' from that of estimating its parameters and their errors, we may be able to deal with objection 3. The question of the best criterion is still open.

The situation is rather characteristic of the effect of a distinctly dangerous influence, as dangerous as a power tool in a child's hand. Criteria previously shelved, <u>partly</u> because of the computation labor, suddenly become computable, but that does not necessarily mean that they are appropriate. In a sober assessment of the computational power given to the statistician he must realize that this merely reopens the discussion on

certain of these criteria which used to be computationally cumbersome. It does not mean that they should be used. The power of the high speed computor will therefore reopen the case for numerous criteria and statistical methods, notably those concerned with searching the data to discover things, discovering unusual features in the data pointing to important exceptions to the law assumed, discovering a regression law, discovering the appropriate metameters. Let us use this power tool judiciously!

REFERENCES

Box, G. E. P. (1949) 'A General Distribution of Likelihood Criteria'. Biometrika, 36, 317.

David, F. N. (1938) 'Tables of the Correlation Coefficient'. Cambridge University Press.

Fieller, E. C. and Hartley, H. O. (1954) 'Sampling with Control Variables'. Biometrika 41, 494.

Fisher, R. A. (1954) 'The Analysis of Variance with Various Binomial Transformations.' Biometrics, 10, 130.

Geary, R. C. (1947) 'The Frequency Distribution of $\sqrt{b_1}$ for Samples of All Sizes Drawn at Random from a Normal Population.' Biometrika 34, 70.

Grubbs, F. E. (1950) 'Sample Criteria for Testing Outlying Observations'. Ann. Math. Stat. 21, 27.

Hammersley, J. M and Morton, K. W. (1955) 'A New Monte Carlo Technique Antithetic Variates'. Proc. Cambridge Phil. Soc. 52, 449.

Hartley, H. O. (1940) 'Testing the Homogeneity of a Set of Variances'. Biometrika 31, p. 249.

Hartley, H. O. (1956) 'A Plan for Programming Analysis of Variance for General Purpose Computers' Biometrics, 12, 110.

Kahn, H. and Marshall, A. W. 'Methods of Reducing Sample Size in Monte Carlo Computations. J. Operations Research Soc. of America 1, 5, 263.

Kempthorne, O. (1953) Query 104. Biometrics 9, 528.

Marshall, A. W. 'The Use of Multi Stage Sampling Schemes in Monte Carlo Computations.' Gainesville Symposium. (1954.)p. 123.

Massey, F. J. (1951) 'The Kolmogorov Smirnov Test for Goodness of Fit'. J.A.S.A. 46, 68.

Trotter, H. F. and Tukey, J. W. 'Conditional Monte Carlo for Normal Samples.' Gainesville Symposium. (1954). p. 64

von Neumann, J., Kent, R. H., Bellinson, H. R. and Hart, B. I. (1941) 'Distribution of the Ratio of the Mean Square Successive Difference to the Variance.' Ann. Math. Stat. 12, 367.